

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局(43) 国际公布日:
2002年5月23日(23.05.02)

PCT

(10) 国际公布号:
WO 02/40528 A1

(51) 国际分类号: C07K 14/435, C12N 15/12, 15/85, 5/16

(21) 国际申请号: PCT/CN01/01506

(22) 国际申请日: 2001年10月26日(26.10.01)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
00125859.1 2000年10月26日(26.10.00) CN

(71) 申请人(对除美国以外的所有指定国): 中国科学院上海生物化学研究所(SHANGHAI INSTITUTE OF BIOCHEMISTRY, CHINESE ACADEMY OF SCIENCES) [CN/CN]; 中国上海市岳阳路 320 号, Shanghai 200031 (CN).

(72) 发明人, 及

(75) 发明人/申请人(仅对美国): 陆长德(LU, Changde) [CN/CN]; 中国上海市岳阳路 320 号, Shanghai 200031 (CN). 黄君霆(HUANG, Junting) [CN/CN]; 中国江苏省镇江市中国农科院蚕研所, Jiangsu 212018 (CN). 赵均(ZHAO, Yum) [CN/CN]; 张峰(ZHANG, Feng) [CN/CN]; 陈秀(CHEN, Xiu) [CN/CN]; 中国上海市岳阳路 320 号, Shanghai (CN).

(74) 代理人: 上海专利商标事务所(SHANGHAI PATENT & TRADEMARK LAW OFFICE); 中国上海市桂平路435号, Shanghai 200233 (CN).

(81) 指定国(国家): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW

(84) 指定国(地区): ARIPO专利(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), 欧亚专利(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲专利(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI专利(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

本国际公布:
— 包括国际检索报告。

所引用双字母代码和其它缩写符号, 请参考刊登在每期 PCT公报期刊起始的“代码及缩写符号简要说明”。

(54) Title: A METHOD OF PRODUCING NON-NATURAL SILK BY BOMBYX MORI

(54) 发明名称: 利用家蚕生产非天然蚕丝的方法

(57) Abstract: The invention is related to the fields of protein engineering and DNA recombination technology. Specially, the invention is related to a method which modify natural silk (including modifying the heavy chain and light chain of the fibroin) by DNA recombination technology and protein engineering technology to obtain natural silk which performance has been improved and to get Bombyx mori which can produce the said natural silk. The invention is also related to the method producing the said natural silk.

(57) 摘要

本发明涉及蛋白质工程和 DNA 重组技术领域, 更具体地, 本发明涉及用 DNA 重组技术和蛋白质工程技术对蚕丝(包括对丝素蛋白重链和轻链)进行改造的方法、由此获得的性能改进的蚕丝、以及产生所述蚕丝的家蚕, 以及生产所述蚕丝的方法。

WO 02/40528 A1

利用家蚕生产非天然蚕丝的方法

技术领域

5 本发明涉及蛋白质工程和 DNA 重组技术领域, 更具体地, 本发明涉及用 DNA 重组技术和蛋白质工程技术对蚕丝进行改造的方法、由此获得的性能改进的蚕丝、以及产生所述蚕丝的家蚕, 以及生产所述蚕丝的方法。

发明背景

10 家蚕是一种能吐丝的经济昆虫, 丝素蛋白基因可在五龄家蚕的后部丝腺中特异而高效表达, 早就受到研究者的广泛注意。有关转基因家蚕的研究也受到国内外学者的重视。李振刚利用天蚕丝基因启动子与家蚕丝素蛋白启动子结构相似的特性, 将完整的天蚕丝基因导入家蚕内, 试图使它得到表达。Marshall 直接地提出将外源基因置于丝素蛋白启动子的调控下, 利用家蚕丝腺做为表达外源基因的生物反应器(Marshall, A. *Nature Biotechnology*, 1998, 16:530-533)。

15 蜘蛛丝质地轻、富有弹性、强度高, 具有广泛应用前景。例如, 可用作光学显微镜上的指示针、外科手术缝合线和制作降落伞、防弹衣等; 可作为保护层包裹在航天器表面以防流星或陨石的碰撞。长期以来, 各国科学家对蜘蛛丝的研究一直很有兴趣, 试图大量获取和利用蜘蛛丝。

20 以园蛛属蜘蛛为例, 它有 7 种不同的丝腺分泌丝, 其中以大囊状腺分泌的拖牵丝性能最优。其独特的强度和弹性的结合使得它的断裂能远高于其他的天然纤维和人工纤维。1960 年, Warwicker 利用 X-ray 研究 *N. madagascarensis* 的拖牵丝时发现结晶区由反平行的皱褶状的 β -折叠组成, 晶体与无规区相比是占多数的。Lewis 等最先根据拖牵丝部分水解获取的短肽序列设计寡核苷酸探针筛选大囊状腺(major ampullate gland) cDNA 文库获得成功。在最近的几篇综述(Hinman MB, Jones JA, Lewis RV. *Trends*
25 *Biotechnol.* 2000, 18(9):374-379; Hayashi CY, Lewis RV. *Science*, 287:1477-1479.; Gosline JM, Guerette PA, Ortlepp CS, et al. *The Journal of Experimental Biology*, 1999, 202(23): 3295-3303)详细地叙述了多种蜘蛛丝蛋白的研究进展, 并且对它们的蛋白质一级结构进行了比较和分析, 有一些肽序列多次出现, 被认为是蜘蛛丝蛋白的基本单元, 即 GPGXX, GGX, 聚丙氨酸, GPX 和间隔区(spacer)。

30 有关拖牵丝结构的研究中, 最有说服力的是 Simmons 等的研究, 他们的研究表明, 90% 的 Ala 在折叠结构中, 而 70% 的 Gly 在无定型区内。使用固相 $^1\text{H-NMR}$ 研究未定向、定向和压缩状态的纤维。结果表明拖牵丝纤维的结晶区中有两类丙氨酸富集区, 一类是高度定向的; 而另一类是定向较差的, 并且也不是紧密堆积的。总之, 现在关于蜘蛛丝强度及弹性的认识是: 强度来自于结晶区, 在一级结构上讲是“聚丙氨酸”这样的序列; 而弹性与
35 β -转角是相关的, 而在一级结构上讲是 GPGXX 这样的序列。

有关拖牵丝的表达工作已在大肠杆菌和酵母中进行。由于拖牵丝基因中含有大量的重复序列, PCR 克隆有困难, 研究者们多倾向于采用人工合成、逐步加倍的方式获得基因。

在人工合成时还兼顾了表达宿主的密码子偏好性。另外，由于大囊状腺蛋白中究竟有几种蛋白及它们之间的比例如何仍未知，所以研究者倾向于同时表达已知的两种蛋白，并将这两种蛋白按一定的比例进行组合。在对表达的产物研究中，多采用了固定金属亲和层析的方法纯化蛋白。对蛋白产物的鉴定中也多集中于 Western 检测、氨基酸组成分析和 N-端测序的方法，也有人采用了 MS 的方法。对于产物二级结构的分析则仅限于 CD 分析。在表达研究中遇到的困难，特别突出的是表达蛋白的成纤维化问题。

鉴于家蚕是一种能吐丝的经济昆虫，人们自然地关注能否用转基因的手段在家蚕中表达蜘蛛拖牵丝。转基因蚕研究已具有较高的水准，它的主要研究内容有 3 个方面：1) 将外源基因导入蚕卵的方法。人们可以将外源 DNA 重组到 NPV (核多角体病毒) 中，利用 NPV 将外源基因导入家蚕细胞、幼虫中。其中 AcNPV 虽可在家蚕中复制，但受感染的幼虫不表现出核多角体的症状，因而也不失为一种有效的方法。其他方法中，起初人们使用得较多的是显微注射的方法。近年来，基因枪、电穿孔都获得了发展。电穿孔、基因枪都在不同的实验室中采用，证明是有效的。2) 报告基因的选择和后代的筛选方法。CAT 基因、GFP (绿色荧光蛋白) 基因和 *neo*^R (新霉素抗性基因) 等都已成功地应用于转基因蚕的研究。3) 将外源基因携带入基因组的方式。早期的研究中多使用了随机重组的方式。使用转座子也是一种重要的方法。然而，现在越来越多的研究者关注外源基因在基因组中的重组位置。一则，担心外源基因的随机重组导致功能基因的失活，二则，认为重组的位置同外源基因的表达效率有关。关于这一点，近年来张峰和 Yamao 都取得了一定的进展。张峰等利用家蚕丝素蛋白的重链基因的 5'-和 3'-端序列为同源片段，介导 IE 启动子调控下的 GFP 基因对重链基因的剔除。Southern 杂交结果表明了同源重组事件的发生，即利用同源重组定向改进蚕丝性能是可能的。Yamao 等则针对丝素蛋白的轻链进行同源重组的尝试，在轻链中融合入 GFP 基因，在基因、蛋白水平上都找到了相应的证据(Yamao M, Katayama N, Nakazawa H, et al. Gene & Development, 1999, 13(5): 511~516)。

由于蚕丝是一种重要的织物原料，因此本领域迫切需要开发新性能的蚕丝，及其生产方法。

发明概述

在本发明的第一方面，提供了一种蚕丝，它基本上由重组的家蚕丝素蛋白构成，所述的重组的家蚕丝素蛋白具有修饰或改变的重链和/或轻链，

其中所述的修饰或改变的重链在重组家蚕丝素蛋白的重链的 N 端非结晶区和 C 端非结晶区中保留天然家蚕丝素蛋白中相应的天然序列，并且在重组家蚕丝素蛋白的中间结晶区中插入了选自下组的外源蛋白或其片段：绿色荧光蛋白、蜘蛛拖牵丝蛋白，和/或重组家蚕丝素蛋白的部分或全部的中间结晶区被选自下组的外源蛋白或其片段置换：蜘蛛拖牵丝蛋白、修饰的蚕丝素蛋白结晶区、蜘蛛丝-蚕丝混合丝蛋白、蜘蛛拖牵丝蛋白-绿色荧光蛋白的融合蛋白；

其中所述的修饰或改变的轻链包括在轻链的外显子中插入选自下组的外源蛋白或

其片段：绿色荧光蛋白、蜘蛛拖牵丝蛋白。

较佳地，所述的重组家蚕丝素蛋白的全部的中间结晶区被蜘蛛拖牵丝蛋白-绿色荧光蛋白的融合蛋白所置换。

在本发明的一个实施例中，所述的蜘蛛拖牵丝蛋白的长度为 800 氨基酸或更多，
5 并且具有 8 个重复单元或更多。

在本发明的另一实施例中，所述的蜘蛛拖牵丝蛋白-绿色荧光蛋白融合蛋白具有 SEQ ID NO: 1 所示的氨基酸序列。

在本发明的另一实施例中，所述的重组家蚕丝素蛋白的轻链具有选自下组的修饰：

10 (1)在外显子 2 和 3 之间插入选自下组的外源蛋白或其片段：绿色荧光蛋白、蜘蛛拖牵丝蛋白；

(2)在外显子 6 和 7 之间插入选自下组的外源蛋白或其片段：绿色荧光蛋白、蜘蛛拖牵丝蛋白；

在本发明的第二方面，提供了一种家蚕丝素蛋白，所述的家蚕丝素蛋白具有修饰
15 或改变的重链和/或轻链，

其中所述的修饰或改变的重链是在重组家蚕丝素蛋白的重链的 N 端非结晶区和 C 端非结晶区中保留天然家蚕丝素蛋白中相应的天然序列，并且在重组家蚕丝素蛋白的中间结晶区中插入了选自下组的外源蛋白或其片段：绿色荧光蛋白，蜘蛛拖牵丝蛋白，

和/或重组家蚕丝素蛋白的部分或全部的中间结晶区被选自下组的外源蛋白或其片段置换：蜘蛛拖牵丝蛋白、修饰的蚕丝素蛋白结晶区、蜘蛛丝-蚕丝混合丝蛋白、蜘蛛拖牵丝蛋白-绿色荧光蛋白的融合蛋白；
20

其中所述的修饰或改变的轻链包括在轻链的外显子中插入选自下组的外源蛋白或其片段：绿色荧光蛋白、蜘蛛拖牵丝蛋白。

在本发明的第三方面，提供了一种核酸分子，它编码本发明上述的家蚕丝素蛋白的修饰或改变的重链和/或轻链。此外，还提供了一种载体，它含有本发明上述的核酸分子。还提供了一种宿主细胞，还宿主细胞含有所述的载体。
25

在本发明的第四方面，提供了一种转基因家蚕，它合成、分泌、吐出本发明上述的蚕丝。

在本发明的第五方面，提供了一种产生本发明蚕丝的方法，培养权利要求 10 所述的转基因家蚕使其分泌蚕丝，以获得蚕丝。
30

在本发明第六方面，提供了一种产生转基因家蚕的方法，包括步骤：

(a)提供一融合基因，该基因含有融合的蜘蛛拖牵丝基因和报告基因，并且在融合基因的 5'端和 3'端侧分别接家蚕丝素蛋白重链 5'端和 3'端序列，

(b)将步骤(a)中的融合基因导入家蚕的蚕卵；

35 (c)使步骤(b)中的蚕卵孵化、发育和结茧；

(d)鉴别基因组中整合有融合基因的家蚕，从而获得转基因家蚕。

较佳地，所述的报告基因是绿色荧光蛋白；在步骤(b)中通过电穿孔法将融合基因

附图说明

附图中, 图 1 显示了人工合成的蜘蛛拖牵丝基因单元。

图 2 显示了 GFP-蜘蛛拖牵丝融合蛋白原核表达质粒 pEGS1000。

图 3 显示了原核表达质粒中 GFP-拖牵丝融合基因的序列。其中, T7 启动子、SD 序列 用下划线标记; 翻译起始密码子 ATG 和终止密码子 TAA 阴影标记。GFP 基因的序列以斜体标记, GFP 基因前后两端的 PstI(CTGCAG)方框标记; 拖牵丝基因中含有 4 个合成的基因单元, BglII (AGATCT), BamHI (GGATCC), HindIII(AAGCTT), NheI(GCTAGC) 及 BglII 和 BamHI 连接形成的 GGATCT 以下划线标记。

图 4 显示了 GFP-蜘蛛拖牵丝融合蛋白真核表达质粒。

图 5 显示了融合蛋白的氨基酸序列。其中, 阴影部分表示拖牵丝序列, 其中的多聚丙氨酸序列黑体和下划线标记。

图 6 显示了对大肠杆菌表达的 GFP-拖牵丝融合蛋白的 SDS-PAGE 和 Western 杂交分析。其中各泳道分别是: M, 蛋白分子量标记; 1, 诱导后 8 h; 2, 诱导后 4 h; 3, 诱导后 2 h; 4, 诱导后 1 h; C, BL21(DE3)空菌。

图 7 显示了对质粒 pEGS1000 在 BL21(DE3)中受 IPTG 诱导后不同时间的荧光强度分析结果。

图 8 显示了融合蛋白荧光性质的分析结果。其中图 8A 为激发光谱, 图 8B 为发射光谱。

图 9 是质粒 p53GS 的核心部分的示意图。其中, GFP-拖牵丝融合基因和重链基因的接合处对应的氨基酸序列在图中标记出。融合基因编码的序列由斜体表示, 重链基因编码的序列由黑体表示。

图 10 是显示了重链修饰的转基因家蚕的"亮茧"和对照蚕茧。其中, T, "亮茧"; C, 对照。

图 11 对基因组家蚕中同源重组事件的 PCR 鉴定示意图。其中, PCR 产物(11,12)的出现表示同源重组事件的发生, PCR 产物(11,13)的出现则表示未被改变的家蚕丝素蛋白重链基因仍存在。

图 12 是同源重组事件的 PCR 鉴定电泳图谱。图 12A 中, 各泳道是: M, DNA 分子量标记; 1, GFP 基因的 PCR 产物; 2, 11-13 PCR 产物; 3, 11-12 PCR 产物。图 12B 中, 各泳道是: M, DNA 分子量标记; 1, 11-13 PCR 产物; 2, 11-12 PCR 产物。

图 13 是同源重组事件的 Southern 杂交鉴定示意图。其中, GFP 基因约 0.7 Kb, 拖牵丝基因约 2.4 Kb。家蚕基因组 DNA 经 HindIII 酶解后与两种探针进行杂交。探针 1 是重链基因上游 HindIII--XhoI 约 1.3 Kb 的序列, 探针 2 是 GFP 基因。

图 14 是同源重组事件的 Southern 杂交鉴定结果。其中图 14A 为重链探针杂交结果, 图 14B 图为 GFP 探针杂交结果。泳道 C 为对照的 54A 家蚕; 泳道 T 为转基因家蚕。

图 15 显示了与拖牵丝基因的点杂交结果。其中点 C 为对照的 54A 家蚕；点 1, 2 为转基因家蚕。

图 16 显示了家蚕丝素蛋白轻链基因的部分序列。图中，标明了用于扩增 L2 的引物 1、2 的位置，L2 的扩增产物约为 950bp；扩增 L3 的引物 3、4 的位置，L3 的扩增产物约为 770bp；在转基因家蚕的鉴定中，设计了一对引物(引物 7、8)用于鉴定未重组事件的存在，它们的位置标在其中，扩增产物约为 400bp。引物上的限制性内切位点 HindIII(AAGCTT)、PstI(CTGCAG)、BglII(AGATCT)和 EcoRI(GAATTC)用黑体标记。

图 17 显示了 GFP 基因与轻链基因外显子 2、3 融合之后的部分基因序列。图中 GFP 基因及其编码的氨基酸为黑体部分，GFP 基因两端的 PstI (CTGCAG)和 BglII (AGATCT)用方框标出。转基因家蚕鉴定中用于检测同源重组事件的引物位置用下划线标记。

图 18 为质粒 p23G 的组装示意图。

图 19 为质粒 p23G 的酶解鉴定图谱。图中，各泳道分别为：M, 分子量标准；1, BglII/EcoRI；2, BglII/PstI；3, PstI/HindIII；4, EcoRI/HindIII；5, EcoRI；6, BglII；7, PstI；8, HindIII。

图 20 显示了在轻链外显子 2 和 3 之间插入 GFP 的转基因家蚕的“亮茧”和对照蚕茧。其中，T, “亮茧”；C, 对照。

图 21 是对转基因蚕 GFP-L23 进行 PCR 鉴定示意图。

图 22 显示了家蚕丝素蛋白轻链基因的部分序列。图中标明了用于扩增 L6 的引物 1、2 的位置，L6 的扩增产物约为 1200bp；扩增 L7 的引物 3、4 的位置，L7 的扩增产物约为 490bp；在转基因家蚕的鉴定中，设计了一对引物(引物 7、8)用于鉴定未重组事件的存在，它们的位置标在图中，扩增产物约为 500bp。引物上的限制性内切位点 HindIII(AAGCTT)、PstI(CTGCAG)、BglII(AGATCT)和 EcoRI(GAATTC)用黑体标记。

图 23 显示了 GFP 基因与轻链基因外显子 6、7 融合之后的部分基因序列。图中 GFP 基因及其编码的氨基酸为黑体部分，GFP 基因两端的 PstI(CTGCAG)和 BglII (AGATCT)用方框标出。用于检测同源重组事件的引物(引物 9 和引物 10)位置为下划线标记。

图 24 为质粒 p67G 的组装示意图。

图 25 为质粒 p67G 的酶解鉴定图谱。图中，各泳道分别为：M, 分子量标准；1, BglII/EcoRI；2, BglII/PstI；3, PstI/HindIII；4, EcoRI/HindIII；5, EcoRI；6, BglII；7, PstI；8, HindIII。

图 26 是对转基因蚕 GFP-L67 进行 PCR 鉴定示意图。

图 27 是对转基因蚕 GFP-L67 进行 PCR 鉴定的电泳图谱。图中，各泳道分别为：M, DNA 分子量标记；1, GFP 基因；2, 5'未重组事件；3, 5'同源重组事件。

图 28 是质粒 p53GS 的构建示意图。

定义

如本文所用，术语“蚕丝”指家蚕吐出的丝。蚕丝的化学组成主要有丝素和丝胶

两种蛋白质成份, 还含有极少量的蜡物质、碳水化合物、色素和无机物。一根茧丝由两根平行的单丝组成它们借丝胶粘合而成。每根单丝中轴为丝素, 外覆一层胶质称为丝胶。

如本文所用, 术语“丝素蛋白”是蚕丝的主要组成成份, 丝素蛋白聚集成为纤维的骨架。

如本文所用, 术语“拖牵丝”指蜘蛛大囊状腺分泌的丝, 是一种性能优良的蛋白质纤维, 具有良好的强度和伸度。蜘蛛利用拖牵丝做为固定蜘蛛网的骨架。

如本文所用, 术语“拖牵丝的基本结构单元”指拖牵丝中一段肽的特异的氨基酸序列, 它通常包括 GPGXX, GGX, 聚丙氨酸(区), GPX 和间隔区等。其中聚丙氨酸区起结晶作用; GPGXX、GGX 和 GPX 属于起转折作用的区域; 而间隔区在聚丙氨酸区和转折区之间起过渡作用。拖牵丝的基本结构单元可以是天然的或人工合成的, 通常具有至少 1 个, 较佳地 2—100 个聚丙氨酸区, 至少 1 个(较佳地 2—100 个)选自 GPGXX、GGX 和 GPX 的转折区, 以及至少 1 个(较佳地 2—100 个)间隔区。其长度通常为 200—6400 个氨基酸或更长, 较佳地为 400—3200 个氨基酸。

如本文所用, 术语“报告基因”指一种基因, 它所编码的蛋白产物便于检测, 在转基因研究中用于筛选和检测转基因个体。

如本文所用, 术语“绿色荧光蛋白(Green Fluorescent Protein, GFP)”指从多管水母(*Aequorea victoria*)中发现的, 蓝光的激发可使其产生绿色荧光的蛋白。基于 GFP 的发光特性, 它在转基因动物研究中已被广泛使用。GFP 已在昆虫细胞中得到表达, 而拖牵丝在昆虫细胞中的表达尚未见报道。天然的或改造过的 GFP 都可用于本发明。

如本文所用, 术语“亮蚕”指在紫外线照射下, 会发出荧光的蚕茧。在本发明中, 因为转基因家蚕所分泌的蚕丝中含有 GFP, 因此紫外线照射下会发出荧光。利用这一性状可以快速筛选出转基因家蚕。

发明详述

蜘蛛有多种腺体可分泌丝, 其中以大囊状腺(major ampullate gland)分泌的拖牵丝(dragline)性能最优。拖牵丝中至少存在两种蛋白, 被分别称为 MaSp1 和 MaSp2, Lewis 等克隆了编码这两种拖牵丝蛋白的 cDNA, 提供了研究拖牵丝结构和功能的模型(1 Xu M, Lewis R V., Proc. Natl. Acad. Sci. USA, 1990, 87(18): 7120~7124; Hinman M B, Lewis R V. J Biol Chem, 1992, 267(27): 19320~19324)。MaSp1 中具有典型的多聚丙氨酸序列, Jelinski 等研究了拖牵丝蛋白的结构, 结果表明 90%的丙氨酸残基在折叠结构中, 而 70%的甘氨酸残基在无定型区内; 拖牵丝纤维的结晶区中有两个丙氨酸富集区, 一个是高度定向的, 而另一个是定向较差的, 并且也不是紧密堆积的。拖牵丝蛋白中的多聚丙氨酸对拖牵丝具有良好强度具有重要作用。后来的研究中也认为多聚丙氨酸和拖牵丝的强度有关。关于蜘蛛丝的异源表达已经有了很多报道, 拖牵丝蛋白在大肠杆菌(*E. coli*)和甲醇酵母(*Pichia pastoris*)中都已获得表达, 表达产物经圆二色性(CD)检测表明具有 β -折叠结构, 和天然的拖牵丝蛋白一致(Prince J T, McCrath K P, DiGrolano C M, et al.

Biochemistry, 1995, 34(34): 10879~10885; Fahnestock S R, Irwin S L. *Appl Microbiol Biotechnol*, 1997, 47(1): 23~32; Fahnestock S R, Bedzyk L A. *Appl Microbiol Biotechnol*, 1997, 47(1): 33~39)。

在本发明的一个实施例中，合成了编码蜘蛛拖牵丝蛋白部分序列的基因单元片段，加倍至 1200bp 后与 GFP 基因融合，在大肠杆菌和昆虫细胞中表达，表达产物具有同 GFP 一样的荧光性质。而后，将长度为 2400 bp 的人工拖牵丝基因和 GFP 基因融合，两侧装配家蚕丝素蛋白重链基因 5'和 3'端序列，用电穿孔方法导入蚕卵。蚕卵孵化、发育和结茧后，紫外灯下筛选到了“亮茧”。对“亮茧”蚕的后代进行鉴定，结果表明 GFP 基因和蜘蛛拖牵丝基因存在于家蚕基因组中并可遗传；PCR 和 Southern 杂交证明转基因蚕的同源染色体中一条发生了预期的同源重组事件；“亮茧”的丝蛋白在 ELISA 实验中可以与 GFP 多克隆抗体反应；对蚕丝的性能也进行了测定，其强度和伸度都比对照组家蚕有所降低。由此，得到了 GFP-蜘蛛拖牵丝融合基因转基因蚕，它能作为进一步用蛋白质工程改造蚕丝性能的中间体。

在本发明另一实施例中，分别将 GFP 基因插入在家蚕丝素蛋白轻链基因外显子 2、3 之间，和外显子 6、7 之间，其后代中都可观察到“亮茧”的出现。基因鉴定说明 GFP 基因存在于家蚕基因组中并可遗传；PCR 证实存在预期的同源重组事件。实现了对丝素蛋白轻链基因进行的基因改造。分别得到了杂合体的转基因家蚕 GFP-L23 和 GFP-L67。

适用于本发明的外源蛋白（或其片段）可以来自不同的物种(如各种蜘蛛)，也可以来自同一物种的不同品系，可以是天然的，也可以是人工合成的。代表性的外源蛋白（或其片段）包括(但不限于)：绿色荧光蛋白、蜘蛛拖牵丝蛋白、修饰的蚕丝素蛋白结晶区、蜘蛛丝-蚕丝混合丝蛋白、蜘蛛拖牵丝蛋白-绿色荧光蛋白的融合蛋白。一种特别优选的外源蛋白是由人工合成的 DNA 序列所编码的拖牵丝。

下面结合具体实施例，进一步阐述本发明。应理解，这些实施例仅用于说明本发明而并不用于限制本发明的范围。下列实施例中未注明具体条件的实验方法，按照常规条件如 Sambrook 等人，分子克隆：实验室手册(New York: Cold Spring Harbor Laboratory Press, 1989)中所述的条件，或按照制造厂商所建议的条件。

实施例

30 实施例 1

GFP-蜘蛛拖牵丝蛋白基因在大肠杆菌和家蚕细胞中的表达

在本实施例中，根据蜘蛛(*Nephila clavipes*)拖牵丝(dragline)蛋白一级结构，人工合成编码蜘蛛拖牵丝蛋白部分序列的基因单元片段，此单元片段中含有的基本结构单元有：3 个聚丙氨酸，2 个 GPGXX，14 个 GGX，和间隔区。并利用两端的 *Bam*HI 和 *Bgl*II 加倍至 1200 bp，与 GFP 基因融合后，在大肠杆菌和昆虫细胞中进行表达。融合蛋白的分子量经 Western 检测确证为 60 KD，与预测值相符。对融合蛋白的荧光性质

进行了分析, 结果表明它的荧光性质与 GFP 一致, 即激发峰为 490 nm, 发射峰为 510 nm。同时研究了变性剂、加热等条件下融合蛋白荧光性质的稳定性。

实施例 1A

5 人工蜘蛛拖牵丝基因的合成及表达质粒的构建

1. 人工合成蜘蛛拖牵丝基因单元及 GFP 基因的克隆

蜘蛛拖牵丝基因单元采用了全合成的方式, 共合成了 8 条寡核苷酸片段, 两两配对形成双链, 每一对双链片段两端各有 6 核苷酸的粘性末端互连后, 连接入 pUC19 的 *Pst*I 和 *Sac*I 之间, 其序列经测序确证与设计相符合。基因单元内含有编码聚丙氨酸、GGX 和 GPGXX 等几种拖牵丝蛋白基本结构单元(module)。在基因单元的 5'-端还设计了 *Bgl*III, 与 3'-端的 *Bam*HI 可用于基因单元的加倍(见图 1)。在基因单元的最前端设计了 *Pst*I 位点, 这个 *Pst*I 位点用于与 GFP 基因融合。

根据 pGreen Lantern 中 GFP 基因的序列设计并合成了一对 PCR 引物(引物 A 和引物 B) 用于扩增 GFP 基因, 其中, 划线部分为 *Eco*RI (GAATTC)和 *Pst*I (CTGCAG):

15 引物 A: 5'-ATAGAATTCCTGCAGGCGGCATGAGCAAGGGCGAGGAA-3'(SEQ ID NO: 2)

引物 B: 5'-AAACTGCAGGAATTCCTTGTACAGCTCGTCCATGCC-3' (SEQ ID NO: 3)

在引物设计时, 使得拖牵丝基因的序列能够与 GFP 基因的阅读框架相吻合。pGreen Lantern 中的 GFP 基因是经过改造的, 将 65 位的 Ser 定点突变为 Thr 使得它的激发峰为 490 nm, 发射峰为 510 nm; 而有别于野生型的 GFP(激发峰为 395nm, 发射峰为 510 nm), 且荧光强度增加(Heim R, Cubitt A B, Tsien R Y. Nature, 1995, 373(6516): 663~664)。

25 2. GFP-蜘蛛拖牵丝基因原核表达质粒的构建

对 pUC19 进行了改造, 去除了 *Eco*RI 和 *Nar*I 之间的一段序列, 包括其中的一个 *Pvu*II 位点, 然后又去除了 *Hind*III 和剩余的一个 *Pvu*II 之间的 *LacZ* 启动子的序列。在此基础上继续构建表达质粒。引入了 T₇启动子的序列, 下游又引入了 SD 序列, 在 *Hind*III 之后加入了翻译起始密码子 ATG; 为了方便纯化, 融合基因的 3'端的翻译终止密码子 TAA 前还加入了编码 6 个连续的 His 的序列, TAA 后侧引入了一个位点 *Nhe*I。最后根据的文献报道(Brosius J, Dull T J, Sleeter D, et al., J Mol Biol, 1981, 148(2): 107~127), 利用 PCR 扩增了大肠杆菌 *rrnB* 基因的终止子序列置于质粒中。将 GFP-拖牵丝融合基因克隆入质粒, 最终形成原核表达质粒 pEGS1000(见图 2)。其中从 T₇启动子到终止密码子 TAA 的 DNA 序列见图 3。此外, 因为拖牵丝基因中有多个 *Pst*I 位点, 为了克隆的方便, 在质粒构建中可使用合成的蚕丝基因做为过渡。

合成片段引入 T₇启动子的接头: 划线部分为 T₇启动子

5'-TAATACGACTCACTATAGGGAAGTCTCGAGACTTCCCTA-3' (SEQ ID NO: 4)

3'-ACGTATTATGCTGAGTGATATCCCTTCAGAGCTCTGAAGGGATTCTGA-5' (SEQ ID NO: 5)

合成片段引入 SD 序列的接头: 划线部分为 SD 序列

5 5'-AGCTTTTATATTTGGAGGTAATCATATGGCTGCA-3' (SEQ ID NO: 6)

3-'AATATAAACCTCCATTAGTATACCG-5' (SEQ ID NO: 7)

合成片段在表达载体的末端引入 TAA 的接头: 划线部分为 TAA, 它的前面是编码 6 个连续 His 的序列

10 5'-CATCACCATCACCATCACTAATGCTAGCG-3' (SEQ ID NO: 8)

3'-TCGAGTAGTGGTAGTGGTAGTGATTACGATCGCTTAA-5' (SEQ ID NO: 9)

克隆 *rrnB* 终止子序列的 PCR 引物: 划线部分为 *EcoRI* 和 *AatII* 位点

引物 C: 5'-TAGAATTCAGAAGTGAAACGCCG-3' (SEQ ID NO: 10)

15 引物 D: 5'-TAAGACGTCGATGGCTTGTAGATATGAC-3' (SEQ ID NO: 11)

3. GFP-蜘蛛拖牵丝基因真核表达质粒的构建

在质粒 pGL₃(Promega 公司)的多克隆位点的 *BglII* 和 *HindIII* 之间克隆入家蚕核多角体病毒即刻早期蛋白基因(IE)启动子 pIE, 而后将 pEGS1000 中的 *HindIII* 和 *NheI* 之间的 GFP-拖牵丝融合基因克隆入 *HindIII* 和 *XbaI* 位点之间, 形成真核表达质粒 pIEGS(见图 4.)。

从 GFP-拖牵丝原核、真核表达质粒的构建来看, 两者所表达的基因是完全一致的, DNA 序列见图 3。表达产物为 661 个氨基酸残基组成的蛋白质, 理论分子量经 DNastar 软件计算约为 60 KD, 其序列如图 5 所示。

25

实施例 1B

GFP-拖牵丝融合蛋白在大肠杆菌中的表达及鉴定

将实施例 1A 中构建的原核表达质粒 pEGS1000, 转化入感受态的大肠杆菌 BL21 (DE3)中, IPTG 诱导表达。在诱导后的 1、2、4、8 小时之后, 取一定体积的菌液离心, 30 进行 SDS-PAGE (10%)分离和 Western 检测。

通过 Western 检测, 表明融合蛋白的分子量为 60 KD, 与预期值的相符; 同时用 Western 检测的结果也可看出表达产物随时间的增加在积累(见图 6)。

荧光强度分析时, 收集各个时间的相同数量的菌体, 离心, 经液氮反复冻融后用 1ml 水悬浮, 再离心后取上清进行测定。将激发光定在 470 nm, 扫描不同时期菌体内 35 容物的发射光谱。荧光强度的测定分析与 Western 分析的结果是一致的(见图 7)。上述两项实验中, 测定各个时期菌液的 OD₆₀₀, 调整所分析菌液的体积, 使得分析的菌体量一致。

GFP-蜘蛛拖牵丝融合蛋白的荧光性质分析

取 IPTG 诱导后 8 h 的菌液, 进行融合蛋白的荧光性质分析。将发射光定在 510 nm, 进行扫描发现激发峰位于 490 nm; 将激发光定在 470 nm, 进行扫描发现发射峰位于 510 nm。在分析发射光谱时, 因激发峰(490 nm)与发射峰(510 nm)相差小, 为避免干扰将
5 激发光设定在 470 nm。这个结果与实验中使用的 GFP 的荧光性质是相同的, 说明融合蛋白未影响 GFP 的荧光性质, 融合蛋白具有同 GFP 一致的荧光性质。

GFP-蜘蛛拖牵丝融合蛋白的存在形式分析

取 IPTG 诱导后 8 h 的菌液, 10,000 转/分离心, 取上清为 A; 菌体经液氮裂解之后, 再经 10,000 转/分离心, 上清为 B; 剩余的部分经水重悬为 C。这三个部分经荧
10 光光谱分析后发现, 有荧光的为 B。这表明: 融合表达产物存在于胞内且是可溶性的。

GFP-蜘蛛拖牵丝融合蛋白的荧光性质的热稳定性分析

用水加热的方式溶解丝胶, 是各种从蚕茧中去除丝胶的方法中较温和的一种方法。取大肠杆菌经 IPTG 诱导 8 h 后的表达产物进行融合蛋白荧光性质的热稳定性分析。先获取若干菌液, 离心后用液氮破裂, 用 100 μ l 水溶解, 离心, 取上清。另取几只离
5 心管加入 1 ml 水, 预先放置在所需温度的水浴中 10 min。待测定时, 再将裂解菌液各 20 μ l 加入各个离心管中, 放置一定时间进行测定。其中放置在室温中的一只做实验对照。分别在 100 $^{\circ}$ C 加热 5 min、10 min 和 20 min, 结果发现融合产物的发光特性在 5 min 时就开始丧失。降低温度, 50 $^{\circ}$ C 加热 30 min, 表达产物的荧光性质仍存在, 但荧光强度有所减弱(未示出)。
10

实施例 1C

GFP-蜘蛛拖牵丝融合基因在 BmN 细胞中的瞬时表达

(1) 用 lipofectin 转染 BmN 细胞

在 35mm 直径的培养皿中, 加入 Bm-N 细胞 (Grace, T.D.C., Nature, 1967, 216:613)
15 1 ml (2.0×10^5 个/ml), 再加 3 ml 含 10% 的 FBS 的 TC-100 细胞培养液, 27 $^{\circ}$ C 培养 12~24 h 至细胞达 50~80% 的贴壁率。取 2~25 μ l 的 lipfectin 溶液(GIBCO BRL 公司)加入 100 μ l 的无血清培养基中, 另取 1~2 μ g 要转化的 DNA 加入 100 μ l 的无血清培养基中, 放 30 min; 再将两溶液混合继续放置 10~15 min。同时将培养板中的细胞培养基去除, 用无血清培养基清洗 2 次。加入无血清培养基 800 μ l, 并加入上述用 lipfectin 处理的 DNA
10 混合液。27 $^{\circ}$ C 培养 6 h 后, 移去培养基, 换 10% FBS 以及含抗生素的 TC-100 培养基。继续培养 24~96 h 观察或检测外源基因的表达。

结果: 用 lipofectin 将实施例 1A 中构建的 pIEGS 转导入 BmN 细胞, 并在转导后的不同时间将细胞置于荧光显微镜下观察。发现带有绿色荧光的细胞从转染后 24 小时出现, 36 小时后数目逐渐增加, 荧光强度也随之增强, 直至 90 小时仍有大量的荧光
15 细胞存在。收集破碎细胞后, 测定荧光的激发峰、发射峰与 GFP 一致(结果与图 8 类似)。证明 GFP 在与拖牵丝蛋白融合后能在家蚕细胞中表达且荧光性质没有改变。为继续研究 GFP-拖牵丝的转基因蚕和使用 GFP 筛选转基因蚕提供了依据。

实施例 2

GFP-蜘蛛拖牵丝基因转基因家蚕

在本实施例中, 根据蜘蛛(*Nephila clavipes*)拖牵丝(dragline)蛋白结构, 人工合成
5 编码蜘蛛拖牵丝蛋白部分序列的基因片段单元, 并利用两端的 *Bam*HI 和 *Bgl*II 加倍至
约 2400bp 和 GFP 基因融合, 融合基因的两侧装配家蚕丝素蛋白重链基因 5'和 3'端序
列之后, 利用电穿孔方法导入蚕卵中。蚕卵孵化、发育和结茧后, 用紫外灯检查, 在约
5400 个茧中有 73 个“亮茧”, 对其后代进行基因和蛋白质水平上的鉴定。PCR 和 Southern
杂交的结果表明, GFP 基因存在于家蚕基因组 DNA 中, 发生了预期的同源重组事件;
10 点杂交表明人工合成的拖牵丝基因存在于家蚕基因组中; 茧蛋白在 ELISA 反应中可以
与 GFP 的多克隆抗体反应。上述结果说明 GFP-拖牵丝融合基因已通过同源重组整合
入家蚕基因组, 表达产物能参与泌丝, “亮茧”这一表型能用于筛选转基因蚕。

材料:

15 用于引入 *Xho*I 位点的接头 I: 划线部分为 *Xho*I

5'-AGCTTCTCGAGCTGCA-3' (SEQ ID NO: 12)

3'-AGAGCTCG-5' (SEQ ID NO: 13)

用于引入 *Mlu*I 位点的接头 II: 划线部分为 *Mlu*I

5'-CACGCGTG-3' (SEQ ID NO: 14)

20 3'TCGAGTGCGCACTTAA-5' (SEQ ID NO: 15)

用于扩增重链基因 5'-端序列的 PCR 引物: 划线部分为 *Xho*I(CTCGAG)和
*Pst*I(CTGCAG)

引物 5: 5'-ATTTTCATAACCCTCGAG-3' (SEQ ID NO: 16)

引物 6: 5'-ACCGACTGCAGCACTAG-3'(SEQ ID NO: 17)

25 用于扩增重链基因 3'-端序列的 PCR 引物: 划线部分为 *Mlu*I(ACGCGT)和
*Eco*RI(GAATTC)

引物 7: 5'-TAAACGCGTGACTTTGAAACTGGAAG-3' (SEQ ID NO: 18)

引物 8: 5'-TTGTCTTCTAGGAATTC-3' (SEQ ID NO: 19)

鉴定同源重组事件的引物:

30 引物 11: 5'-TTTTCATAACCTCGAGGTAGC-3' (SEQ ID NO: 20)

引物 12: 5'-CATGGCACAGGGAGCTTTC-3' (SEQ ID NO: 21)

鉴定未重组事件的引物:

引物 11: 5'-TTTTCATAACCTCGAGGTAGC-3' (SEQ ID NO: 22)

引物 13: 5'-CTCCTGCACCGACTGCAG-3' (SEQ ID NO: 23)

35 GFP 基因鉴定引物

引物 14: 5'-ATGAGCAAGGGCGAGGA-3' (SEQ ID NO: 24)

引物 15: 5'-TTGTACAGCTCGTCCATG-3' (SEQ ID NO: 25)

扩增重链基因探针的 PCR 引物:

引物 16: 5'-TATAAGCTTGTGTACAAAAGT-3' (SEQ ID NO: 26)

引物 17: 5'-TGCTACCTCGAGGTTATG-3' (SEQ ID NO: 27)

扩增丝素蛋白重链基因 3'端非翻译区的 PCR 引物:

5 引物 18: 5'-GCCAGTGAATTCCTAGAAGA-3' (SEQ ID NO: 28)

引物 19: 5'-CCAGGACGAAGTAAGAAAC-3' (SEQ ID NO: 29)

方法

1. 外源基因转入蚕卵的方法

- 10 p53GS 经 *ScaI* 酶解、线性化处理后, 酚/氯仿(1:1)、氯仿抽提, 乙醇沉淀后溶于双蒸水中。将刚产下的家蚕受精卵, 放入电穿孔杯中(400~500 粒/杯), 加线性化 DNA 的溶液(0.2 mg/ml)混匀, 冰浴 5 分钟。在电穿孔仪(BIO-RAD gene pluser II)中, 以 500 V、25 pF 和 200 Ω 的条件下电击, 取出蚕卵, 晾干, 浸酸处理。

15 2. PCR 反应条件

使用 GIBCO BRL 公司的快速抽提试剂 DNAzol™ 从蚕蛾及蚁蚕中抽提基因组 DNA。操作方法参照商品说明书改进如下: 样品在玻璃均浆器中反复碾磨, 15,000 转/分离心, 等体积酚: 氯仿抽提, 15,000 转/分离心, 水相加 1/2 体积乙醇沉淀, 95% 乙醇洗涤沉淀, 凉干后, 用水溶解, 供 PCR 反应使用。

- 20 PCR 反应的反应体系为: 模板为 100 ng 家蚕 DNA; 引物为 20 pmol, 总反应体积为 50 μ l。反应程序基本上为 94.0℃变性 7 min; 94.0℃变性 30 s, 52.0℃退火 30 s, 72.0℃延伸 30 s, 共 35 个循环; 72.0℃延伸 7 min; 4.0℃保温。

3. Southern 分析

25 3.1 家蚕 *Bombyx mori* 基因组 DNA 的抽提

在家蚕五龄的第三天, 解剖取出后部丝腺体约 1 g, 放入研钵中加入液氮, 待其完全蒸发后, 用力碾碎, 最后加入 3~4 ml DNA 抽提缓冲液(10 mmol/L Tris-HCl (pH 8.0), 100 mmol/L EDTA (pH 8.0), 0.5% SDS), 碾均匀。然后加入 RNase 至终浓度 20 μ g/ml, 37℃, 1 h。再加入蛋白酶 K 至终浓度 100 μ g/ml, 边加边打匀, 50℃, 3 h。

- 30 用等体积酚抽提 3 次。最后, 加入 1/10 体积 NaAc(3 mol/L, pH 5.2), 2 体积乙醇沉淀, 4℃放置过夜, 挑取絮状沉淀(由于丝腺中有较多的丝蛋白存在, 沉淀时会与 DNA 一同沉淀下来, 使用较长的时间沉淀后, DNA 的絮状沉淀会漂浮在溶液上层, 而蛋白形成的颗粒状沉淀处于溶液的下方, 进一步起到分离的作用), 70%乙醇洗涤。风干, 水或 TE 溶解。

35

3.2 基因组 DNA 的酶解和电泳

用限制性内切酶 *HindIII* 酶解, 37℃ 2 小时。酶解处理后的 DNA 经等体积酚: 氯

仿抽提、乙醇沉淀后,加水溶解,50℃处理 5min,然后加上样缓冲液上样,待 DNA 样品静置 5min 后,进行电泳。电泳时,前 5min 使用较高电压,使 DNA 样品快速进胶,而后在不超过 1V/cm 条件下进行电泳。

5 3.3 杂交

转膜后,在预杂交液(6 × SSC, 0.5% SDS, 5 × Denhardt 溶液, 100 μg/ml 鲑精 DNA)中于 68℃预杂交 2 h,加入热变性处理的探针于 68℃杂交 8 h。

室温条件下于 2 × SSC、0.5% SDS 及 2 × SSC、0.1% SDS 中依次漂洗 5 min, 0.1 × SSC 0.1% SDS 中于 68℃漂洗 30 min。放入 X-光片夹内,进行放射自显影。

10

4. 丝蛋白的溶解及 ELISA 实验

蚕茧放入清水中煮沸 30 min, 3 次,清水中的成份为丝胶蛋白。留下的蚕丝放入 6 mol/L LiBr 中溶解, 65℃过夜,其后对水透析,而透析后的产品主要为丝素蛋白,可能会含有少量丝胶成份。

15

4.1 试剂

包被液(pH 9.5): Na_2CO_3 0.8 g, NaHCO_3 1.46 g, 蒸馏水 500 ml 溶解

洗涤液(pH 7.4): Tris 6 g, 1 mol/L HCl 37 ml, 25% Tween-20 5 ml, 蒸馏水 2500 ml 溶解

20

底物缓冲液: $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ 8.72 g, 柠檬酸 2.72 g, 蒸馏水 500 ml 溶解

抗体稀释液(pH 7.5): NaCl 4 g, KH_2PO_4 0.1 g, KCl 0.1 g, $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ 1.45 g, 25% Tween-20 2 ml, 蒸馏水 500 ml 溶解

封板液: PBS 配制的 15%的小牛血清溶液

25

4.2 操作

➤ 用包被液将抗原稀释成 5 μg/ml, 在酶标板中每孔加入 100 μl 上述抗原溶液, 37℃, 1 h

➤ 除去酶标板内的包被液。用洗涤液洗涤 3 次, 每次 5 min

➤ 每孔加入封板液 100 μl, 37℃, 1 h

30

➤ 除去封板液。用洗涤液洗涤 3 次, 每次 5 min

➤ 用抗体稀释液(含有 1%的小牛血清)适当稀释一抗, 每孔加入 100 μl, 37℃, 1 h

➤ 弃去抗体溶液, 用洗涤液洗涤 3 次, 每次 5 min

➤ 用抗体稀释液(含有 1%的小牛血清)适当稀释二抗, 每孔加入 100 μl, 37℃, 1 h

➤ 弃去抗体溶液, 用洗涤液洗涤 3 次, 每次 5 min

35

➤ 加入底物(四甲基联苯胺 2 mg 溶于 5 ml 底物 buffer)10 μl, 5 μl H_2O_2 , 每孔 50 μl。37℃反应适当时间, 2 mol/L H_2SO_4 50 μl 终止反应

➤ 测定 OD_{450}

实施例 2A

同源重组质粒的构建

1. 人工合成蜘蛛拖牵丝基因单元

- 5 蜘蛛拖牵丝采用了全合成的方式，共合成了 8 条寡核苷酸片段，两两配对形成双链，每一对双链片段两端各有 6 核苷酸的粘性末端互连后，连接入 pUC19 的 *Pst*I 和 *Sac*I 之间。在基因单元的 5'-端还设计了 *Bgl*II，与 3'-端的 *Bam*HI 可用于基因单元的加倍。基因单元序列见第一部分，在基因单元的最前端有 *Pst*I 位点可用于 GFP 基因的融合，使得拖牵丝基因与 GFP 基因的阅读框架吻合。天然的蚕丝是对光不稳定的，原因在于丝蛋白中含有不少的 Tyr。因此在人工合成的蜘蛛拖牵丝基因中，有意将 Tyr 更换为 Thr。

2. 对家蚕进行成体实验的质粒 p53GS 的构建

- 15 蚕丝素蛋白 H 链基因由两个外显子和一个内含子组成。外显子 I 所编码的序列中有典型的信号肽序列，可使得后部丝腺中合成的丝蛋白分泌出去。外显子 II 则编码丝素蛋白重链的重复区，在蛋白的 C-端存在 3 个 Cys 残基，其中两个形成链间二硫键，另一个与轻链形成链间二硫键。设计了一对引物扩增基因 5'端 *Xho*I 至 *Pst*I 间的约 950 bp 的序列。*Xho*I 位于内含子中，而 *Pst*I 位于外显子 II 的前端，其后的序列是丝素蛋白重链基因的重复区域。为了克隆的方便使用了其中 *Pvu*II 至 *Pst*I 的约 700 bp 的序列充当同源区。如果发生正确的同源重组，GFP-拖牵丝基因将在丝素蛋白重链基因的启动子调控下表达，并且使用丝素蛋白重链的 N-端信号肽序列分泌。用于扩增家蚕丝素蛋白重链基因 3'端下游同源序列的引物序列为：正向引物 5'-TAAACGCGTGACTTTGAAAC TGGAAG-3'(引物 7, SEQ ID NO: 18)，其中划线部分为引入的 *Mlu*I 位点，反向引物 5'-TTGTCTTCTAG GAATTC-3'(引物 8, SEQ ID NO: 19)，正向引物中的 *Mlu*I 是为了克隆的便利而在引物中引入的。这一段基因在丝素蛋白重链基因重复序列之后，发生同源重组之后不改变原基因的末端序列，GFP-拖牵丝基因只是替代丝素蛋白重链基因的重复序列，这样基因产物的 C-端的 3 个 Cys 残基将不受影响，可与轻链共价结合。由于基因的下游不受改变，基因转录后的后加工(如，加 polyA 等)应不受影响。

- 25 分别改造 pUC19，使其原有的 *Sph*I 被替代为 *Xho*I，称 pUX；在 *Sac*I 和 *Eco*RI 之间加入位点 *Mlu*I，称 pUM。丝素蛋白重链 5'端序列克隆于 pUX 的 *Xho*I 和 *Pst*I 之间，以后截去 *Xho*I 至 *Pvu*II 的序列，仅保留 *Pvu*II 至 *Pst*I 约 700 bp 的序列，丝素蛋白重链 3'端序列克隆于 pUM 的 *Mlu*I 和 *Eco*RI 之间，5'端序列和 3'端序列拼接在一起形成过渡质粒 p52。丝素蛋白重链 3'端非翻译区的序列克隆在 pUC18 中，称为 pUC100。p52 和 pUC100 经 *Aat*II 和 *Eco*RI 酶解后，将丝素蛋白重链 3'端非翻译区的序列也插入 p52 中，形成过渡质粒 p53。拖牵丝基因单元以 *Bgl*II 和 *Bam*HI 加倍后直至 2.4 Kb 后，克隆于 p53 的 *Pst*I 和 *Sac*I 之间。从 pGreen Lantern 质粒中扩增的 GFP 基因插入在 *Pst*I 位点，利用 GFP 基因中的 *Eco*RI 位点确定插入的方向。最后构成了 p53GS 质粒(详细的质粒

构建过程见附图 28)。如实施例 1 中提及, GFP 基因与拖牵丝基因的融合具有共同的阅读框架, 在本实施例中, GFP-拖牵丝融合基因与克隆的两段重链基因也是融合的, 其接头部位的融合设计如图 9 所示。

5 实施例 2B

GFP-蜘蛛拖牵丝融合基因转基因家蚕的筛选

1. 转基因处理的蚕当代(G_0)性状表现

用 *ScaI* 酶解质粒 p53GS, 使其线性化, 用电穿孔方法处理了约一万颗蚕卵, 分 5 天收蚁 6700 头左右, 孵化率为 67%, 而一般没有经过处理的蚕卵, 二日孵化率都达 90% 以上。蚕饲养到上簇、结茧后第五天, 用紫外灯观察蚕茧, 结果在转基因处理的 5436 个蚕茧中, 有 73 个亮茧, 而在对照组中没有“亮茧”出现(见图 10)。

2. 家蚕基因组 DNA 的 PCR 检测(G_0 、 G_1 和 G_2)

对 G_0 、 G_1 代中“亮茧”中的蚕蛾进行 PCR 检测, 都检出了 GFP 基因, 并依据 PCR 实验的结果保留检出明显 PCR 产物的蛾区。 G_2 代中共获得 35 个蛾区, 抽提蚕蛾的 DNA, 用 PCR 检测出 10 对含有 GFP 基因。35 个蛾区的蚕卵待孵化后取出约 1/4 的蚁蚕抽提 DNA, 用 PCR 又检测出 13 个蛾区中含有 GFP 基因。鉴定到的阳性蛾区内进行“亮茧”蚕的交配、制种。

20 实施例 2C

转基因家蚕的基因和蛋白质分析

1. 同源重组事件的 PCR 检测

根据 Tsujimoto 和 Suzuki(1979)提供的家蚕丝素蛋白重链基因的 5'-端序列信息, 设计了 3 条 PCR 引物用于检测转基因家蚕。设计了引物 11 为公用的上游引物 5'-TTTTCATAACCTCGAGGTAGC-3'(SEQ ID NO: 20), 它含盖了 *XhoI* 酶切位点, 位于 5'同源重组片段(*PvuII*--*PstI*)的外侧。下游引物则设计了两条, 引物 12: 5'-CATGGCACAGGGAGCTTTC-3'(SEQ ID NO: 21), 这条引物位于 GFP 基因内侧, 只有发生了同源重组事件后, 才可能出现这个 PCR 产物(1.1 Kb); 有时同源重组事件仅在一对同源染色体中的一条发生, 为了区别同源重组事件是在一条染色体上发生还是在另一对染色体上发生, 设计了引物 13: 5'-CTCCTGCACCGACTGCAG-3'(SEQ ID NO: 23), 这条引物含盖了一个 *PstI* 位点, 其余序列代表的是丝素蛋白重链基因的序列, 只要家蚕的一条染色体未发生改变, 那么这一 PCR 产物(0.95 Kb)就会出现。

针对 GFP 基因检测呈阳性的转基因家蚕的后代, 又进行了进一步的同源重组事件的 PCR 鉴定。结果如图 12A 所示, 外源基因 GFP 是存在的, 因为表明发生同源重组事件的 PCR 产物(11,12)也出现了。在图 12B 中看得更清楚, 它比 PCR 产物(11,13)大一些。PCR 产物(11,13)的出现, 说明仍然有未被改变的染色体存在, 那么鉴定到的是杂合体。

2.同源重组事件的 Southern 杂交鉴定

在有了上述同源重组事件 PCR 的鉴定结果后，进一步进行了杂交鉴定。根据丝素蛋白重链基因的序列，进行 *Hind*III 的酶解，用家蚕丝素蛋白重链基因和 GFP 基因两种探针进行杂交。其中重链基因探针是基因上游调控区中的 *Hind*III-*Xho*I 的 1.3 Kb 的序列(见图 13)。

杂交结果如图 14 所示。结果表明，GFP 基因确实存在和重链基因上游调控区的连锁关系(两种探针杂交都出现了 5.8 Kb 的条带)。GFP 探针杂交还出现的 4 Kb 片段，与重链基因上游调控区是不连锁的，应当属随机的同源重组。关于它的来源，也可能是转基因家蚕间杂交育种而获得的。

3.拖牵丝基因的点杂交实验

在尼龙膜上分别点上了 54A 家蚕的基因组 DNA 和两个转基因家蚕(GFP 基因 PCR 检测呈阳性)的基因组 DNA 样品。利用人工合成的拖牵丝基因片段(约 1.2 Kb)为模板合成探针，进行杂交。这两个样品相对于 54A 是呈阳性的(图 15)。

4.对“亮茧”丝蛋白的分析

用 50℃加热的条件，获取了一些“亮茧”的丝胶成份，测定荧光，结果荧光不存在。接着，利用 GFP 多克隆抗体(Santa Cruz 公司)进行 ELISA 实验。在紫外灯下挑选一些“亮茧”，用 6 mol/L LiBr 溶解、透析后进行 ELISA 实验，发现其中的一些可以明显地与 GFP 多克隆抗体反应。见表 1。

表 1. “亮茧”丝蛋白进行的 ELISA 分析

样品	对照茧	1# “亮茧”	2# “亮茧”	3# “亮茧”	4# “亮茧”
OD ₄₅₀	0.13	0.33	0.28	0.35	0.27
与对照茧的比值	1.00	2.55	2.17	2.72	2.13

从表 1 数据分析表明茧丝蛋白可以与 GFP 抗体反应，但其中丝胶、丝素的成份还未得到分离。因此，进一步分离丝胶、丝素进行分析。用清水煮沸，水中可以溶解丝胶。剩余的成份用 6 mol/L LiBr 溶解、透析，其中主要是丝素成份，但含有少量丝胶。简单分离后的样品进行 ELISA 分析，发现丝胶不与抗体反应，而丝素(混有少量丝胶)可以与抗体反应。这表明，可以与 GFP 抗体反应的是丝素的成份，见表 2。

表 2. “亮茧”丝胶、丝素的 ELISA 分析

样品	空白	54A 丝胶	54A 丝素	“亮茧”丝胶	“亮茧”丝素
OD ₄₅₀	0.25	0.32	0.33	0.31	0.57
与空白的比值	1.00	1.29	1.36	1.27	2.32

茧中的丝蛋白可以与 GFP 抗体反应说明存在 GFP 的特异序列，由于翻译、泌丝

是同源重组、转录和表达的下流事件，可以认为这项结果是对同源重组的有力支持，同时也说明同源重组的正确性。

5. 蚕丝机械强度的测定

- 5 蚕丝经脱胶处理后，在紫外灯下不再发亮。这与实施例 1 表达研究中发现融合蛋白的荧光性质在现有的脱胶方法中会淬灭的结果(未给出具体数据)是相吻合的。脱胶后的蚕丝用来测定了单丝强度，结果如表 3 所示。与 54A 家蚕相比，强度有所减弱。

- 10 对于这样的结果，我们认为，丝纤维是一种蛋白纤维，蛋白单体间的组合方式对纤维的性质有影响。GFP-拖牵丝融合蛋白与丝素蛋白的相互作用如何(多数为杂合体)，很难预测。杂交结果表明发生同源重组的一条染色体 *HindIII* 酶解片段变小了许多，仅 6 Kb。可见 GFP-蜘蛛拖牵丝融合蛋白的分子量较小，而且其中还有一段 GFP(约 230 氨基酸残基)，因此结晶区更短，推测强度会降低。其次 GFP 可能不与蚕丝、蜘蛛丝形成好的结晶，也能使丝的强度降低。

表 3. 转基因家蚕单丝断裂强度、伸度的测定

名称	断裂强度(厘牛顿)	断裂伸长度(%)
对照组	14.5	32.2
A	11.5	24.4
B	11.0	28.0
C	11.5	24.3
D	10.8	22.7
E	11.6	27.6
F	11.8	25.0
G	8.7	20.6
H	9.7	24.0
I	10.5	24.5
J	10.9	19.5

- 15 注:从每个蛾区中取 10 颗蚕茧，每颗蚕茧抽出的丝分为 50 cm 长取 50 段进行测定，取其平均值代表每颗茧的强度及伸度。再将每个蛾区内的 10 颗茧平均，代表蛾区内的强度、伸度。本实验中采用的是较低温度的缂丝方法，使得茧中的蛹得以存活，可继续发育为蚕蛾完成交配、产卵。

20 实施例 3

GFP-家蚕丝素蛋白轻链融合基因转基因蚕 GFP-L23

在本实施例中，利用 PCR 扩增了家蚕丝素蛋白轻链基因的 2 个片段 L2 和 L3，GFP 基因同 L2 中的轻链基因外显子 2 和 L3 中的外显子 3 融合在一起，用电穿孔方法将这

样的基因片段导入蚕卵中。家蚕经孵化、发育和结茧后，用紫外灯检察，在约 8000 颗蚕茧发现有 280 颗“亮茧”出现。选取 G₃ 代中一个蛾区内的 20 头家蚕，取血并进行 PCR 鉴定，其中有 6 个家蚕个体的基因组 DNA 中同时含有 GFP 基因和预想中的同源重组事件。对相应蚕茧中的蛋白质成份进行分析发现：其中 3 个蚕茧的丝素蛋白可与

5 GFP 多克隆抗体反应。

材料

用于扩增 L2 片段的 PCR 引物：

引物 1:5'-ATTAAGCTTCAAATTCTCTAGAAAAACATT-3' (SEQ ID NO: 30)

引物 2:5'-TTACTGCAGATCACTGTATTGATTGATGG-3'(SEQ ID NO: 31)

10 用于扩增 L3 片段的 PCR 引物：

引物 3:5'-TCAACGTTCAAGAGATCTTG-3' (SEQ ID NO: 32)

引物 4:5'-TAGAATTCGGTAATATTGGATGTTTAAGG-3' (SEQ ID NO: 33)

用于扩增 GFP 基因的引物

引物 5:5'-ATACTGCAGAGCAAGGGCGAGGAACT-3'(SEQ ID NO: 34)

15 引物 6:5'-ATTAGATCTCCTTGTACAGCTCGTCCAT-3'(SEQ ID NO: 35)

鉴定未重组事件的引物

引物 7:5'-CGTCACAGTGCATCTAGTG-3'(SEQ ID NO: 36)

引物 8:5'-GCACAGGCATCACCGGG-3'(SEQ ID NO: 37)

鉴定同源重组事件 3'-端引物

20 引物 9:5'-CCTGCTGGAGTTTGTGACC-3'(SEQ ID NO: 38)

引物 10:5'-CAGGGTTGATGACGAGT-3'(SEQ ID NO: 39)

鉴定同源重组事件 5'-端引物

引物 11:5'-GTTTTCGTGAACACGGACGA-3'(SEQ ID NO: 40)

引物 12:5'-CATGGCACAGGGAGCTTTC-3'(SEQ ID NO: 41)

25 鉴定 GFP 基因的引物

引物 13:5'-ATGAGCAAGGGCGAGGA-3'(SEQ ID NO: 24)

引物 14:5'-TTGTACAGCTCGTCCATG-3'(SEQ ID NO: 25)

实施例 3A

30 同源重组质粒的构建

1.两个 L-chain 基因片段及 GFP 基因的克隆设计

依据 Kikuchi 和 Yamaguchi 等文献报道(Kikuchi Y, Mori K, Suzuki S, et al. Gene, 1992, 110(2): 151~158; Yamaguchi K, Kikuchi Y, Takagi T, et al. J Mol Biol, 1989, 210(1): 127~139), 家蚕丝素蛋白轻链基因共约 14 Kb, 有 7 个外显子和 6 个内含子; 编码 262
35 氨基酸残基, 轻链蛋白的分子量约 25 Kd。外显子 3、5、6 上各有一个 Cys, 其中外显子 3 和外显子 5 上的 Cys 形成链内二硫键, 而外显子 6 上的 Cys 则与重链上的一个 Cys 形成链间二硫键。设计了两对 PCR 引物扩增了两个 L-chain 的基因片段, 分别称为 L2、

L3; L2 起始于内含子 1 中, 终止于外显子 2, 两条 PCR 引物上分别引入了 *Hind*III 和 *Pst*I 位点, 用于基因片段的克隆, *Pst*I 还兼顾与 GFP 基因的融合。L3 起始于外显子 3 中, 终止与内含子 3(见图 1); 上游引物中有 *Bgl*II 的位点, 下游引物中引入了 *Eco*RI 的位点, 便于克隆。L2 长约 950 bp, L3 长约 770 bp, 在 L2、L3 之间会在同源重组中被替代的序列有约 1300 bp, 包括外显子 2、3 中的一部分和整个内含子 2。以 pGreen
5 Lantern 质粒为模板扩增 GFP 基因, 引物上分别设计了 *Pst*I 和 *Bgl*II 酶切位点, 便于产物的克隆并使 GFP 基因于 L2 和 L3 基因片段融合。

按照上述的设计, 外显子 2 编码的 28 个氨基酸残基去除了后端的 13 个, 外显子 3 编码的 65 个氨基酸残基中只去除了前端的 19 个残基。在发生同源重组之后的重组的 GFP-轻链融合基因的部分序列在图 2 中显示。GFP 基因两端的 *Pst*I 和 *Bgl*II 用方框
10 标记出, 在这两个位点附近的氨基酸序也在图中显示。

2. 质粒 p23G 的组装

以 100 ng 家蚕基因组 DNA 为模板, 通过 PCR, 扩增出 L2 和 L3。以 pGreen Lantern 为模板, 扩增了 GFP 基因。L2 的 PCR 产物两端有 *Hind*III、*Pst*I 位点, 克隆于 pUC19
15 中, 为 pL2; GFP 基因的两端有 *Pst*I 和 *Bgl*II, 但 pUC19 中没有 *Bgl*II, 克隆于 *Pst*I 和 *Sma*I 之间, 为 pGFP; L3 两端有 *Bgl*II 和 *Eco*RI, 克隆于 *Sma*I 和 *Eco*RI 之间, 为 pL3。而后, 按图 18 流程, 将 L2、L3 和 GFP 基因拼接起来组装成 p23G 供成体实验使用。p23G 经酶解鉴定正确(图 19)。

20 实施例 3B

外源基因的导入和转基因家蚕的初步筛选

1. 外源基因的导入

利用电穿孔处理了约 12500 颗家蚕受精卵, 其中有约 8000 颗孵化(孵化率为 65%), 待家蚕发育、结茧后在紫外灯下进行观察, 发现了一些“亮茧”的出现(见图 20), 约
25 有 280 颗这样的“亮茧”, 占孵化卵的 3.5%。

2. 初步基因鉴定及交配、育种

选取了 23 个“亮茧”, 取出蚕蛾与 54A 蚕蛾进行交配, 其中包括了正交(雌性 54A 蚕蛾×雄性 “亮茧”蚕蛾)12 对和反交(雌性 “亮茧”蚕蛾×雄性 54A 蚕蛾)11 对。产下的卵孵化后, 每个蛾区内取约 1/4 的蚁蚕抽提 DNA 进行 GFP 基因的 PCR 检测。约
30 有一半的蛾区显示出阳性, 其中正交中的 4/12 显阳性, 反交中的 6/11 显阳性。显示阳性的蛾区内进行了自交。以 GFP 基因为标志进行的第一次基因鉴定, 结果表明: 在进行转基因处理的当代, 观察到“亮茧”的出现, 提示 GFP 基因可能已进入了当代家蚕的染色体中, 但这并不意味着 GFP 基因一定进入生殖细胞并可以遗传, 只有通过检测后代才能筛选到 GFP 基因可遗传的蛾区。检测第 2 代蚁蚕时, 仅 1/2 蛾区可检出 GFP
35 基因。由于可能存在一些假阴性, 实际可遗传的比例应还高一些, 但是在第二代进行蛾区的可遗传分析是必要的。在蚁蚕 GFP 基因 PCR 检测阳性的蛾区中进行自交、育种, 其后代将按孟得尔规律遗传。

实施例 3C

转基因家蚕的基因及蛋白质分析

1. 基因鉴定

5 对初步筛选到得转基因家蚕进行进一步的 PCR 鉴定：以确定 GFP 基因是否正确
地插入在外显子 2 和外显子 3 中，同源重组事件发生在同源染色体的一条上还是两条
上，因而进一步基因鉴定的内容包括 5'、3'的同源重组事件，GFP 基因和 3'未重组事
件。3'未重组鉴定中，上游引物设计在内含子 2 中(引物 7)，而下游引物设计在 L3 中(引
物 8)，这两个引物完全是轻链基因中的，如果 PCR 反应呈阳性则表明所鉴定到的个体
10 中仍有未改变的轻链基因。进行重组鉴定的两组 PCR 实验中。5'端的融合 PCR 鉴定中，
一条引物设计在 GFP 基因中(引物 12)，而另一条引物则设计在 L2 片段的上游(引物 11)
这样的 PCR 产物的出现可以说明 5'端的同源重组事件的发生。3'端的融合 PCR 鉴定中，
一条引物设计在 GFP 基因中(引物 9)，而另一条引物则设计在 L3 片段的下游(引物 10)
这样的 PCR 产物的出现可以说明 3'端的同源重组事件的发生。发生了同源重组的轻链
15 的部分序列已列出(见图 17)，5'同源重组事件 PCR 产物约 1150 bp，3'同源重组事件 PCR
产物约 1000 bp，GFP 基因 PCR 产物为 700 bp，未重组事件的 PCR 产物为 400 bp(在
图 16 中有说明)。

上一代中 GFP 基因检测为阳性的家蚕自交后，其后代中选取一些蛾区，用约 1/4
的蚁蚕抽提 DNA 进行 PCR 鉴定。这里进行的是同源重组事件的基因鉴定，发现了一
20 些阳性蛾区。对于来自于正交的后代，检测了 17 个蛾区，共检测出阳性蛾区 7 个；反
交蛾区 10 个，检测出阳性蛾区 1 个。阳性蛾区继续饲养，而阴性蛾区则放弃饲养。

待五龄后，从阳性蛾区 23N-2(23N 表示其亲本交配的方式为雄性 GFP 基因阳性
蚕蛾与雌性 54A 蚕蛾交配，2 表示实验中的检测编号)内取了二十头家蚕取血进行了完
整的 PCR 基因鉴定，结果见表 4。首先所有的 3'未重组事件都检测到了阳性，说明这
25 些个体中都有未发生改变的染色体存在。另外，这项 PCR 鉴定的结果也表明从蚕血抽
提的 DNA 可以进行 PCR 反应。从鉴定结果中，可发现外源基因的重组情况是复杂的，
有 6 个家蚕个体同时检出了 5'和 3'同源重组事件和 GFP 基因，由于前面提到的 3'未重
组事件的出现可判定为杂合体，表中为加阴影的个体，其中 A1 的 PCR 鉴定电泳图谱
见图 8。有仅检出 3'同源重组事件的(如 A3, A6)，也有仅检出 5'同源重组事件的(如 A15,
30 A16)。特别提出的是 A18 和 A20，两侧的同源重组事件都检出了，而 GFP 基因却没有
检出。依据 PCR 引物的设计来看，3'和 5'同源重组事件的 PCR 鉴定中，各有一条引物
设计在 GFP 基因内，因此若同源重组事件检测为阳性，GFP 基因的检测也应为阳性。
GFP 基因的 PCR 阴性可能为假阴性。对于它们的后代还会继续关注、鉴定。以蚕血中
抽提的 DNA 进行的 PCR 使得基因鉴定的结果不再是粗略的蛾区，而是家蚕个体。A1
35 至 A10 为雌性家蚕，A11 至 A20 为雄性家蚕。根据 PCR 检测的结果 A17 同 A1、A17
同 A9 进行了交配、制种。

表 4. 23N-2 蛾区的个体 PCR 检测结果

编号	3'-同源重组事件	3'-未重组事件	GFP 基因	5'-同源重组事件
A1	+	+	+	+
A2	+	+	+	-
A3	+	+	-	-
A4	+	+	+	+
A5	+	+	+	+
A6	+	+	-	-
A7	+	+	+	+
A8	-	+	-	-
A9	+	+	+	+
A10	+	+	-	-
A11	+	+	-	-
A13	+	+	-	-
A14	+	+	-	-
A15	-	+	-	+
A16	-	+	-	+
A17	+	+	+	+
A18	+	+	-	+
A19	-	+	-	-
A20	+	+	-	+

注：选取雌蚕(A1~A10)、雄蚕(A11~A20)各 10 头，在五龄时从尾角放血，利用血中抽提的 DNA 进行 PCR 鉴定。

2.蛋白质的 ELISA 鉴定

5 按实施例 2 中相同的方法，对上述进行 PCR 鉴定的这二十头家蚕相对应的部分蚕茧也进行了 ELISA 分析，结果见表 5。A1、A4 和 A17 在基因鉴定中被确定为杂合体，茧丝蛋白的用 GFP 抗体检测均为阳性。A7 的家蚕经取血后，未能存活结茧，无法鉴定。A5 和 A9 却未能检测出 ELISA 的阳性。虽然基因鉴定它们显示阳性，但由于 GFP 的表达是基因同源重组的下游事件，可能由于重组并不十分精确而导致 GFP 基因未能正确表达。对于 A20，GFP 基因鉴定中为阴性，但它的茧丝蛋白能与 GFP 抗体反应，证实了 GFP 基因 PCR 检测结果有假阴性的猜想。

表 5. 23N-2 蛾区中的蚕茧的 ELISA 鉴定结果

样品	对照茧	A1	A4	A17	A20
OD ₄₅₀	0.13	0.34	0.39	0.46	0.37
与对照茧的比值	1.00	2.63	3.04	3.63	2.86

上述分析中，丝胶、丝素的成份未得到分离。因此，按实施例 2 中相同的方法，进一步分离丝胶、丝素后进行分析。即用清水煮沸可以溶解丝胶，剩余的成份用 6 mol/L LiBr 溶解、透析，其中主要是丝素成份但含有少量丝胶。这样简单分离后的样品进行 ELISA 分析，发现丝胶是不与抗体反应的，而丝素(混有少量丝胶)可以与抗体反应，那么可以与 GFP 抗体反应的是丝素的成份(见表 6)。

表 6. 蚕茧的丝胶、丝素蛋白成份进行 ELISA 鉴定结果

样品	空白	54A 丝胶	54A 丝素	A1 丝胶	A1 丝素
OD ₄₅₀	0.25	0.32	0.33	0.35	0.59
与空白的比值	1.00	1.29	1.36	1.44	2.39

从蛾区 23N-2 中选取一部分五龄家蚕，解剖获得中部丝腺，从中获取蛋白质进行了分析，其 ELISA 结果见表 7。说明融合蛋白在后部丝腺中合成后，进入中部丝腺并参与泌丝最终出现在蚕茧中，与轻链蛋白经过了一样的生理过程。

表 7. 五龄家蚕中部丝腺蛋白进行 ELISA 鉴定结果

样品	对照丝腺	丝腺 1	丝腺 2
OD ₄₅₀	0.17	0.82	0.66
与对照丝腺的比值	1.00	4.94	3.99

实施例 4

15 GFP-家蚕丝素蛋白轻链融合基因转基因蚕 GFP-L67

在本实施例中，利用 PCR 扩增了轻链基因的 2 个片段 L6 和 L7， GFP 基因与 L6 中的外显子 6 和 L7 中的外显子 7 融合在一起，用电穿孔方法将这样的基因片段导入蚕卵中。家蚕经孵化、发育和结茧后，用紫外灯检察，在约 9600 颗茧中有 260 颗“亮茧”的出现。在 G₃ 代进行的基因鉴定中发现了 5 个阳性蛾区，证明了 GFP 基因的存在和证实了预想中的同源重组事件。

材料

用于扩增 L6 片段的 PCR 引物:

- 引物 1: 5'-AATCTGCAGTGATCCAGCTGATTGGTG-3'(划线为 *Pst*I 位点)(SEQ ID NO: 42)
- 25 引物 2: 5'-ATTAAGCTTTCCAAATAAGAATGTAA-3'(划线为 *Hind*III 位点)(SEQ ID NO: 43)

用于扩增 L7 片段的 PCR 引物:

- 引物 3: 5'- AGATCTATATTGCACAAGCAGCCA-3'(划线为 *Bg*III 位点)(SEQ ID NO: 44)
- 30 引物 4: 5'-ATGAATTCCAACAGTACCGAAATCCAT-3'(划线为 *Eco*RI 位点)(SEQ ID

NO: 45)

用于扩增 GFP 基因的 PCR 引物:

引物 5: 5'-ATACTGCAGAGCAAGGGCGAGGAACT-3' (划线为 *Pst*I 位点) (SEQ ID NO: 34)

5 引物 6: 5'-ATTAGATCTCCTTGTACAGCTCGTCCAT-3' (划线为 *Bgl*II 位点) (SEQ ID NO: 35)

鉴定未重组事件的引物:

引物 7: 5'-CAGTCAACAGATATCGCC -3' (SEQ ID NO: 46)

引物 8: 5'-GATAAGCTCATCTGTCCAA-3' (SEQ ID NO: 47)

10 鉴定同源重组事件 5'-端引物:

引物 9: 5'-GCCAGCAGTGACTCTAGG -3' (SEQ ID NO: 48)

引物 10: 5'-CATGGCACAGGGAGCTTTC -3' (SEQ ID NO: 49)

鉴定 GFP 基因的引物:

引物 11: 5'-ATGAGCAAGGGCGAGGA-3' (SEQ ID NO: 24)

15 引物 12: 5'-TTGTACAGCTCGTCCATG-3' (SEQ ID NO: 25)

实施例 4A

同源重组质粒的构建

1. 两个 L-chain 基因片段及 GFP 基因的克隆设计

20 依据 Kikuchi 等文献报道, 家蚕丝素蛋白轻链基因共约 14Kb, 有 7 个外显子和 6 个内含子; 编码 262 氨基酸残基, 轻链蛋白的分子量约 25KD。外显子 3, 5, 6 上有一个 Cys, 其中外显子 3 和外显子 5 上的 Cys 形成链内二硫键, 而外显子 6 上的 Cys 则与重链上的一个 Cys 形成链间二硫键。由于二硫键对于蛋白质空间结构的重要性, 设计同源重组方案时, 都避免改变它的二硫键。

25 设计了两对 PCR 引物扩增了两个 L-chain 的基因片段, 分别称为 L6、L7; L6 起始于内含子 5 中, 终止于外显子 6, 两条 PCR 引物上分别引入了 *Hind*III 和 *Pst*I 位点, 用于基因片段的克隆, *Pst*I 还兼顾与 GFP 基因的融合。L7 起始于外显子 7 中, 终止于 3'-非翻译区(见图 22); 上游引物中有 *Bgl*II 的位点, 下游引物中引入了 *Eco*RI 的位点, 便于克隆。L6 长约 1200 bp, L7 长约 490 bp, 在 L6、L7 之间会在同源重组中被
30 替代的序列有约 650 bp, 包括外显子 6、7 中的一部分和整个内含子 6。以 pGreen Lantern 质粒为模板扩增 GFP 基因, 引物上分别设计了 *Pst*I 和 *Bgl*II 酶切位点, 便于产物的克隆、与 L6、L7 基因片段融合。

按照上述的设计, 外显子 6 编码的 48 个氨基酸残基保留了前端的 45 个, 外显子 7 编码的 31 个氨基酸残基中保留了后端的 11 个残基。在发生同源重组之后的重组的
35 GFP-轻链融合基因的部分序列在图 23 中显示。GFP 基因两端的 *Pst*I 和 *Bgl*II 用方框标记, 在这两个位点附近的氨基酸序列也在图中显示。

2. 质粒 p67G 的组装

以 100 ng 家蚕基因组 DNA 为模板, 通过 PCR 条件, 扩增了 L6 和 L7, 以 pGreen Lantern 为模板, 扩增了 GFP 基因。L6 的 PCR 产物两端有 *Hind*III、*Pst*I 位点, 克隆于 pUC19 中, 为 pL6; GFP 基因的两端有 *Pst*I 和 *Bgl*II, 但 pUC19 中没有 *Bgl*II, 克隆于 *Pst*I 和 *Sma*I 之间, 为 pGFP; L7 两端有 *Bgl*II 和 *Eco*RI, 克隆于 *Sma*I 和 *Eco*RI 之间, 为 pL7。而后, 按图 24 所示程序, 将 L6、L7 和 GFP 基因拼接起来组装成 p67G 供实施例 4B 使用。p67G 经酶解鉴定正确(图 25)。

实施例 4B

外源基因的导入和转基因家蚕的初步筛选

10 利用电穿孔处理了约 14600 颗家蚕受精卵, 其中有约 9600 颗孵化(孵化率为 66%), 待家蚕发育、结茧后在紫外灯下进行观察, 发现了一些“亮茧”的出现, 约有 260 颗这样的“亮茧”, 占孵化卵的 3.0%。

实施例 4C

转基因家蚕的基因鉴定

15 选取了 22 个“亮茧”, 取出蚕蛾与 54A 家蚕进行交配, 其中包括了正交(雌性 54A 蚕蛾×雄性“亮茧”蚕蛾)10 对和反交(雌性“亮茧”蚕蛾×雄性 54A 蚕蛾)12 对。待它们产下卵、孵化后, 取约 1/4 的蚁蚕抽提 DNA 进行 GFP 基因的检测。约有一半的蛾区显示出阳性。其中正交中的 6/10 显阳性, 反交中的 5/12 显阳性。显示阳性的蛾区内进行了自交, 以 GFP 基因为标志进行了第一次基因鉴定。

在进行转基因处理的当代, 观察到了“亮茧”的存在, 这提示 GFP 基因可能进入了当代家蚕的染色体中, 但这并不意味着 GFP 基因一定可以遗传, 只有通过检测后代才能筛选到 GFP 基因可遗传的蛾区。检测后代蚁蚕时, 仅 1/2 蛾区可检出 GFP 基因。

25 对于 GFP 基因 PCR 检测呈阳性的家蚕后代, 希望能进一步检测 GFP 基因是否正确地整合于轻链基因外显子 6 和 7 之间。设计了引物 9 和 10, 引物 10 位于 GFP 基因内侧, 引物 9 位于片段 L6 的外侧, 这个 PCR 产物用于说明 5'同源重组事件的发生。另外, 引物 7 位于 L6 片段内, 而引物 8 位于被预想剔除的内含子 6 之中, 这个 PCR 产物的出现将说明存在未被改变的同源染色体(见图 26)。

30 从 GFP 基因 PCR 检测呈阳性的蛾区内取约 1/4 的蚁蚕进行 3'-端同源重组事件的基因鉴定, 发现了一些蛾区内的蚁蚕 DNA 可扩增出代表同源重组事件的 PCR 产物。亲本为正交的鉴定了 12 个蛾区, 仅一个蛾区呈阳性; 反交鉴定了 14 个蛾区, 4 个蛾区呈阳性。预想中发生了同源重组的轻链的部分序列已列出(见图 23), 从中可以计算出 5'-同源重组事件 PCR 产物为 1400 bp。PCR 鉴定的电泳图谱见图 27。

35 讨论

本发明与 Yamao 的工作都是针对家蚕丝素蛋白轻链基因进行基因打靶。Yamao 等解剖五龄家蚕观察到了能发绿色荧光的丝腺, 利用 PCR 和 Southern 杂交证明了同源

重组事件的发生, 同时在分析茧蛋白时发现了与 GFP 抗体反应的 52 KD 的蛋白质。但不能观察到茧的任何变化。

本发明同 Yamao 的相比较有所不同: 1) 使用了不同的同源重组和融合的方式。在本工作中 GFP 融合于外显子 2、3 中(去除外显子 2 中的 13 个氨基酸残基、去除外显子 3 的 19 个氨基酸残基和内含子 2), 或外显子 6、7 中(去除外显子 6 中的 3 个氨基酸残基、去除外显子 7 的 20 个氨基酸残基和整个内含子 6;), 融合蛋白都利用轻链基因自身的翻译终止密码子 TAA。Yamao 等用重组 AcNPV 攻击幼虫, 重组 AcNPV 中的 GFP 基因放置于轻链基因外显子 7 中, 5'-端与外显子 7 融合, 3'-端利用 GFP 基因自身的 TGA 为翻译的终止密码子, 融合蛋白将丢掉外显子 7 末端的 19 个氨基酸残基。2) 本发明中使用电穿孔的基因转移方式, 外源基因经过酶解处理。受精初期的蚕卵为单细胞, 在单细胞时发生同源重组才可得到能遗传的转基因蚕, 即丝腺细胞和生殖细胞具有相同的基因组。Yamao 的实验则使用 NPV 介导的方式。在五龄的第一天攻击家蚕, NPV 能进入生殖细胞且发生同源重组并遗传给后代。3) 不同的筛选方式。本发明依据出现的新表型“亮茧”进行, 而他们直接对下一代家蚕进行 PCR 鉴定, 工作量较大。

在本发明提及的所有文献都在本申请中引用作为参考, 就如同每一篇文献被单独引用作为参考那样。此外应理解, 在阅读了本发明的上述讲授内容之后, 本领域技术人员可以对本发明作各种改动或修改, 这些等价形式同样落于本申请所附权利要求书所限定的范围。

权 利 要 求 书

1.一种蚕丝，其特征在于，它基本上由重组的家蚕丝素蛋白构成，所述的重组的家蚕丝素蛋白具有修饰或改变的重链和/或轻链，

- 5 其中所述的修饰或改变的重链在重组家蚕丝素蛋白的重链的 N 端非结晶区和 C 端非结晶区中保留天然家蚕丝素蛋白中相应的天然序列，并且在重组家蚕丝素蛋白的中间结晶区中插入了选自下组的外源蛋白或其片段：绿色荧光蛋白、蜘蛛拖牵丝蛋白，和/或重组家蚕丝素蛋白的部分或全部的中间结晶区被选自下组的外源蛋白或其片段置换：蜘蛛拖牵丝蛋白、修饰的蚕丝素蛋白结晶区、蜘蛛丝-蚕丝混合丝蛋白、蜘蛛拖牵丝蛋白-绿色荧光蛋白的融合蛋白；

其中所述的修饰或改变的轻链包括在轻链的外显子中插入选自下组的外源蛋白或其片段：绿色荧光蛋白、蜘蛛拖牵丝蛋白。

2.如权利要求 1 所述的蚕丝，其特征在于，所述的重组家蚕丝素蛋白的全部的中间结晶区被蜘蛛拖牵丝蛋白-绿色荧光蛋白的融合蛋白所置换。

- 15 3.如权利要求 1 所述的蚕丝，其特征在于，所述的蜘蛛拖牵丝蛋白的长度为 800 氨基酸或更多，并且具有 8 个重复单元或更多。

4.如权利要求 2 所述的蚕丝，其特征在于，所述的蜘蛛拖牵丝蛋白-绿色荧光蛋白融合蛋白具有 SEQ ID NO: 1 所示的氨基酸序列。

- 20 5.如权利要求 1 所述的蚕丝，其特征在于，其中所述的重组家蚕丝素蛋白的轻链具有选自下组的修饰：

(1)在外显子 2 和 3 之间插入选自下组的外源蛋白或其片段：绿色荧光蛋白、蜘蛛拖牵丝蛋白；

(2)在外显子 6 和 7 之间插入选自下组的外源蛋白或其片段：绿色荧光蛋白、蜘蛛拖牵丝蛋白；

- 25 6.一种家蚕丝素蛋白，其特征在于，所述的家蚕丝素蛋白具有修饰或改变的重链和/或轻链，

其中所述的修饰或改变的重链是在重组家蚕丝素蛋白的重链的 N 端非结晶区和 C 端非结晶区中保留天然家蚕丝素蛋白中相应的天然序列，并且在重组家蚕丝素蛋白的中间结晶区中插入了选自下组的外源蛋白或其片段：绿色荧光蛋白，蜘蛛拖牵丝蛋白，

- 30 和/或重组家蚕丝素蛋白的部分或全部的中间结晶区被选自下组的外源蛋白或其片段置换：蜘蛛拖牵丝蛋白、修饰的蚕丝素蛋白结晶区、蜘蛛丝-蚕丝混合丝蛋白、蜘蛛拖牵丝蛋白-绿色荧光蛋白的融合蛋白；

其中所述的修饰或改变的轻链包括在轻链的外显子中插入选自下组的外源蛋白或

其片段：绿色荧光蛋白、蜘蛛拖牵丝蛋白。

7.一种核酸分子，其特征在于，它编码权利要求6中所述的家蚕丝素蛋白的修饰或改变的重链和/或轻链。

8.一种载体，其特征在于，它含有权利要求7所述的核酸分子。

5 9.一种宿主细胞，其特征在于，它含有权利要求8所述的载体。

10.一种转基因家蚕，其特征在于，它合成、分泌、吐出权利要求1所述的蚕丝。

11.一种产生权利要求1所述蚕丝的方法，其特征在于，培养权利要求10所述的转基因家蚕使其分泌蚕丝，以获得蚕丝。

12.一种产生转基因家蚕的方法，其特征在于，包括步骤：

0 (a)提供一融合基因，该基因含有融合的蜘蛛拖牵丝基因和报告基因，并且在融合基因的5'端和3'端侧分别接家蚕丝素蛋白重链5'端和3'端序列，

(b)将步骤(a)中的融合基因导入家蚕的蚕卵；

(c)使步骤(b)中的蚕卵孵化、发育和结茧；

(d)鉴别基因组中整合有融合基因的家蚕，从而获得转基因家蚕。

5 13.如权利要求12所述的方法，其特征在于，所述的报告基因是绿色荧光蛋白；在步骤(b)中通过电穿孔法将融合基因导入家蚕的蚕卵；和步骤(d)中鉴别的进行是通过在紫外光下筛选亮茧。

Ala Ala Asp Gly Arg Ser Gly Gly Leu Gly Gly Gln Gly Pro Gly
GCT GCA GAT GGT AGA TCT GGA GGT CTT GGT GGC CAA GGT CCT GGT
CGA CGT CTA CCA TCT AGA CCT CCA GAA CCA CCG GTT CCA GGA CCA
PstI BglII
Gln Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala Gln Gly Gly
CAG GGT GCT GCA GCT GCT GCC GCA GCT GGA GGT GCT CAA GGT GGT
GTC CCA CGA CGT CGA CGA CGG CGT CGA CCT CCA CGA GTT CCA CCA

Thr Gly Gly Leu Gly Ser Gln Gly Ala Gly Arg Gly Gly Gln Gly
ACT GGT GGA CTT GGA TCA CAA GGT GCT GGT AGA GGT GGA CAA GGT
TGA CCA CCT GAA CCT AGT GTT CCA CGA CCA TCT CCA CCT GTT CCA

Ala Ala Ala Ala Ala Ala Val Glu Gly Gly Ala Gly Gln Gly Gly
GCT GCA GCT GCC GCT GCT GTC GAG GGT GGT GCT GGT CAG GGT GGA
CGA CGT CGA CGG CGA CGA CAG CTC CCA CCA CGA CCA GTC CCA CCT

Thr Gly Gly Leu Gly Ser Gln Gly Ala Gly Arg Gly Gly Leu Gly
ACA GGC GGT CTT GGT TCA CAA GGT GCT GGT AGA GGT GGA CTT GGT
TGT CCG CCA GAA CCA AGT GTT CCA CGA CCA TCT CCA CCT GAA CCA

Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Ala Gly Gly Ala
GGT CAG GGA GCT GGT GCT GCT GCT GCA GCT GCT GCT GGT GGT GCA
CCA GTC CCT CGA CCA CGA CGA CGA CGA CGA CGA CCA CCA CGT

Gly Gln Gly Gly Thr Gly Gly Pro Gly Asn Gln Gly Gly Ser Met
GGT CAA GGT GGT ACT GGT GGA CCA GGT AAT CAA GGC GGA TCC ATG
CCA GTT CCA CCA TGA CCA CCT GGT CCA TTA GTT CCG CCT AGG TAC
BamHI

Glu Leu
GAG CTC
CTC GAG
SacI

图 1

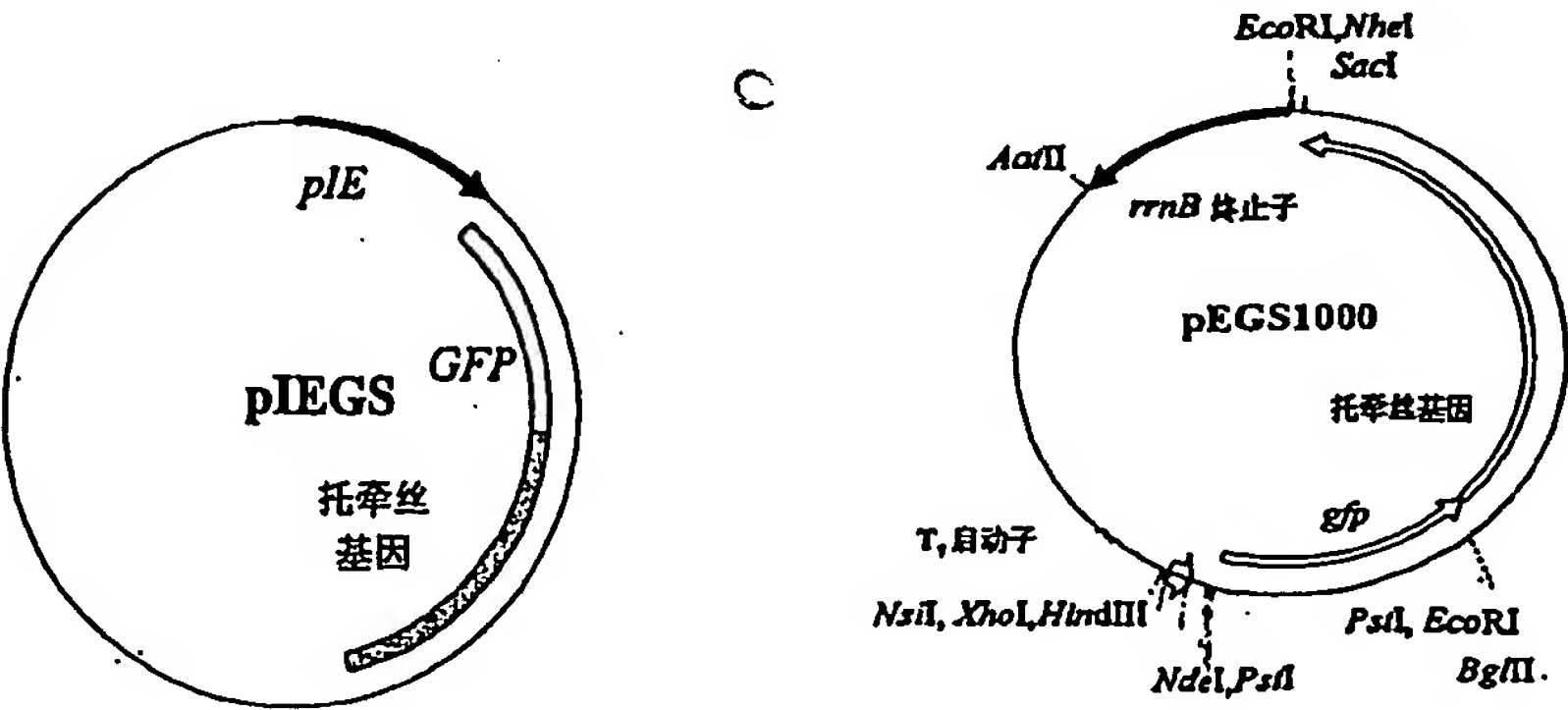


图 2

图 4

ATGCATAATACGACTCACTATAGGGAAGTCTCGAGACTTCCCTAAGCTTTATATTG

GAGGTAATCATATGGCTGCAGGCGGCATGAGCAAGGGCGAGGAACTGTTCACTGGCGTGGTCCCCA
ATTCTCGTGGAAGCTGGATGGCGATGTGAATGGGCACAAATTTTCTGTACGCGGAGAGGGGTGAAGGTG
ATGCCACATACGGAAAGCTCACCCCTGAAATTCATCTGCACCACTGGAAAGCTCCCTGTGCCATGGCC
AACACTGGTCACTACCTTCACCTATGGCGTGCAGTGCTTTTCCAGATACCCAGACCATATGAAGCAGC
ATGACTTTTTCAAGAGCGCCATGCCCCGAGGGCTATGTGCAGGAGAGAACCATCTTTTTCAAAGATGAC
GGGAACTACAAGACCCGCGCTGAAGTCAAGTTCGAAGGTGACACCCTGGTGAATAGAATCGAGCTGA
AGGGCATTGACTTTAAGGAGGATGGAAACATTCTCGGCCACAAGCTGGAATACAACTATAACTCCCAC
AATGTGTACATCATGGCCGACAAGCAAAAGAATGGCATCAAGGTCAACTTCAAGATCAGACACAACAT
TGAGGATGGATCCGTGCAGCTGGCCGACCATTATCAACAGAACTCCAATCGGCGACGGGCCCTGT
GCTCCTCCCAGACAACCATTACCTGTCCACCCAGTCTGCCCTGTCTAAAGATCCCAACGAAAAGAGA
GACCACATGGTCCTGCTGGAGTTTGTGACCGCTGCTGGGATCACACATGGCATGGACGAGCTGTACA
AGGGAATTCCTGCAGATGGTAGATCTGGAGGTCTTGGTGGCCAAGGTCTGGTCAGGGTGCTG
CAGCTGCTGCCGCAGCTGGAGGTGCTCAAGGTGGTACTGGTGGACTTGGATCACAAGGTGCTG
GTAGAGGTGGACAAGGTGCTGCAGCTGCCGCTGCTGTCGAGGGTGGTGGTGGTCAGGGTGGA
ACAGGCGGTCTTGGTTCACAAGGTGCTGGTAGAGGTGGACTTGGTGGTCAGGGAGCTGGTGCT
GCTGCTGCAGCTGCTGCTGGTGGTGCAGGTCAAGGTGGTACTGGTGGACCAGGTAATCAAGGC
GGATCCATGGGATCTGGAGGTCTTGGTGGCCAAGGTCTGGTCAGGGTGCTGCAGCTGCTGCC
GCAGCTGGAGGTGCTCAAGGTGGTACTGGTGGACTTGGATCACAAGGTGCTGGTAGAGGTGGA
CAAGGTGCTGCAGCTGCCGCTGCTGTCGAGGGTGGTGGTGGTCAGGGTGGAACAGGCGGTCTT
GGTTCACAAGGTGCTGGTAGAGGTGGACTTGGTGGTTCAGGGAGCTGGTGGTGGTGGTGGTGGT
GCTGCTGGTGGTGCAGGTCAAGGTGGTACTGGTGGACCAGGTAATCAAGGCGGATCCATGGGA
TCTGGAGGTCTTGGTGGCCAAGGTCTGGTTCAGGGTGCTGCAGCTGCTGCCGCAGCTGGAGGT
GCTCAAGGTGGTACTGGTGGACTTGGATCACAAGGTGCTGGTAGAGGTGGACAAGGTGCTGCA
GCTGCCGCTGCTGTCGAGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
GCTGGTAGAGGTGGACTTGGTGGTTCAGGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
GCAGGTCAAGGTGGTACTGGTGGACCAGGTAATCAAGGCGGATCCATGGGATCTGGAGGTCTT
GGTGGCCAAGGTCTGGTTCAGGGTGCTGCAGCTGCTGCCGCAGCTGGAGGTGCTCAAGGTGG
TACTGGTGGACTTGGATCACAAGGTGCTGGTAGAGGTGGACAAGGTGCTGCAGCTGCCGCTGC
TGTCGAGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
GTGGACTTGGTGGTTCAGGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
GTGGTACTGGTGGACCAGGTAATCAAGGCGGATCCATGGGAGCTCATCACCATCACCATCACTT
AATGCTAGCG

冬 3

MAAGGMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTL
 VTTFTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIEL
 KGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGD
 GPVLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITHGMDLEYKGIPADGRSGGLGGQGP
 QGAAAAAAGGAQGGTGGLGSQAGRGGQAAAAAAVEGGAGQGGTGGLGSQAGRGGGLGG
 QGAGAAAAAAGGAGQGGTGGPNGQGGSMGSGGLGGQGPQAAAAAAGGAQGGTGGLGS
 QGAGRGGQAAAAAAVEGGAGQGGTGGLGSQAGRGGGLGGQAGAAAAAAGGAGQGGTG
 GPGNQGGSMSGGLGGQGPQAAAAAAGGAQGGTGGLGSQAGRGGQAAAAAAVEGGA
 GQGGTGGLGSQAGRGGGLGGQAGAAAAAAGGAGQGGTGGPNGQGGSMGSGGLGGQGPQ
 GAAAAAAGGAQGGTGGLGSQAGRGGQAAAAAAVEGGAGQGGTGGLGSQAGRGGGLGGQ
 GAGAAAAAAGGAGQGGTGGPNGQGGSMGAHHHHHH

图 5

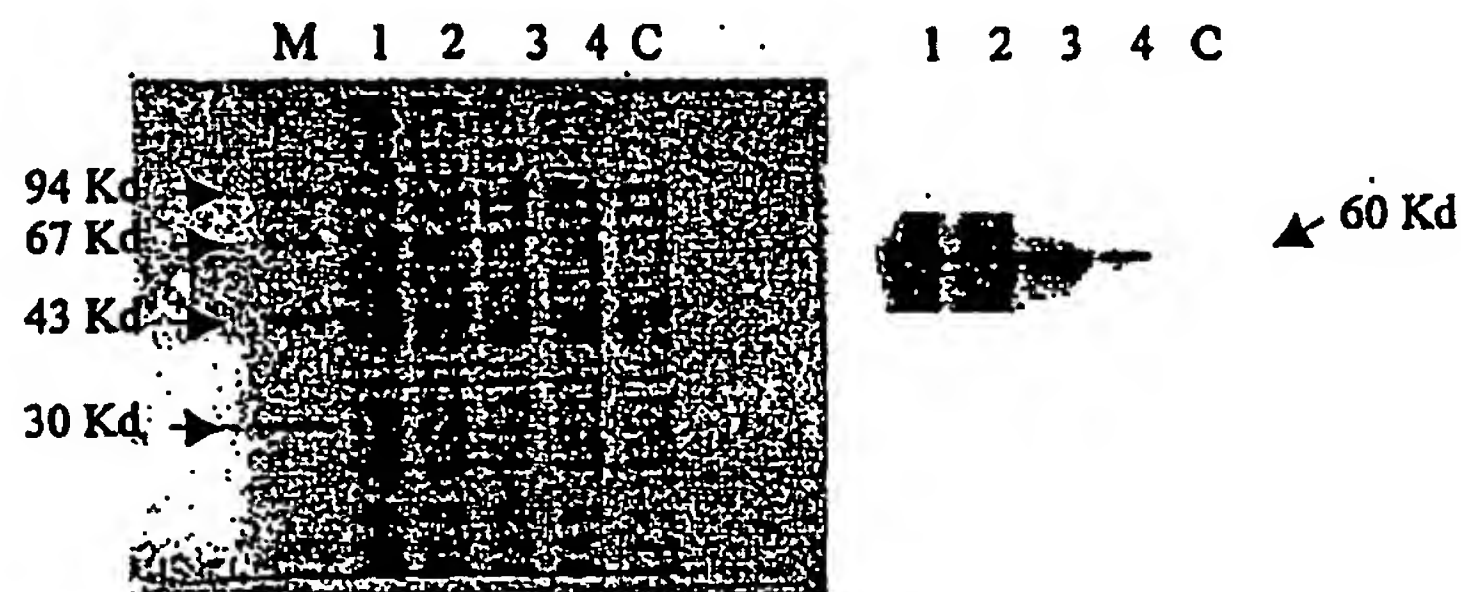


图 6

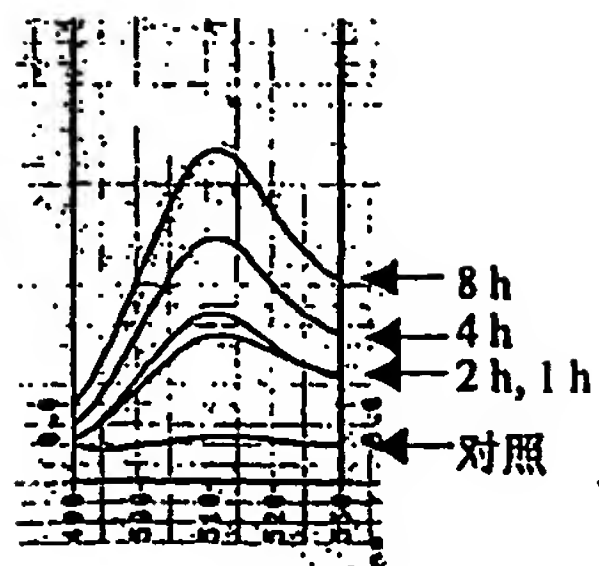


图 7

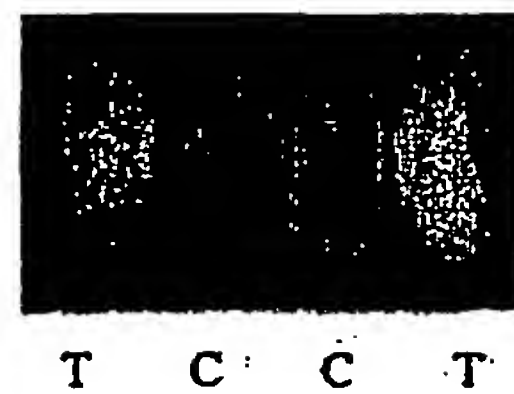


图 10

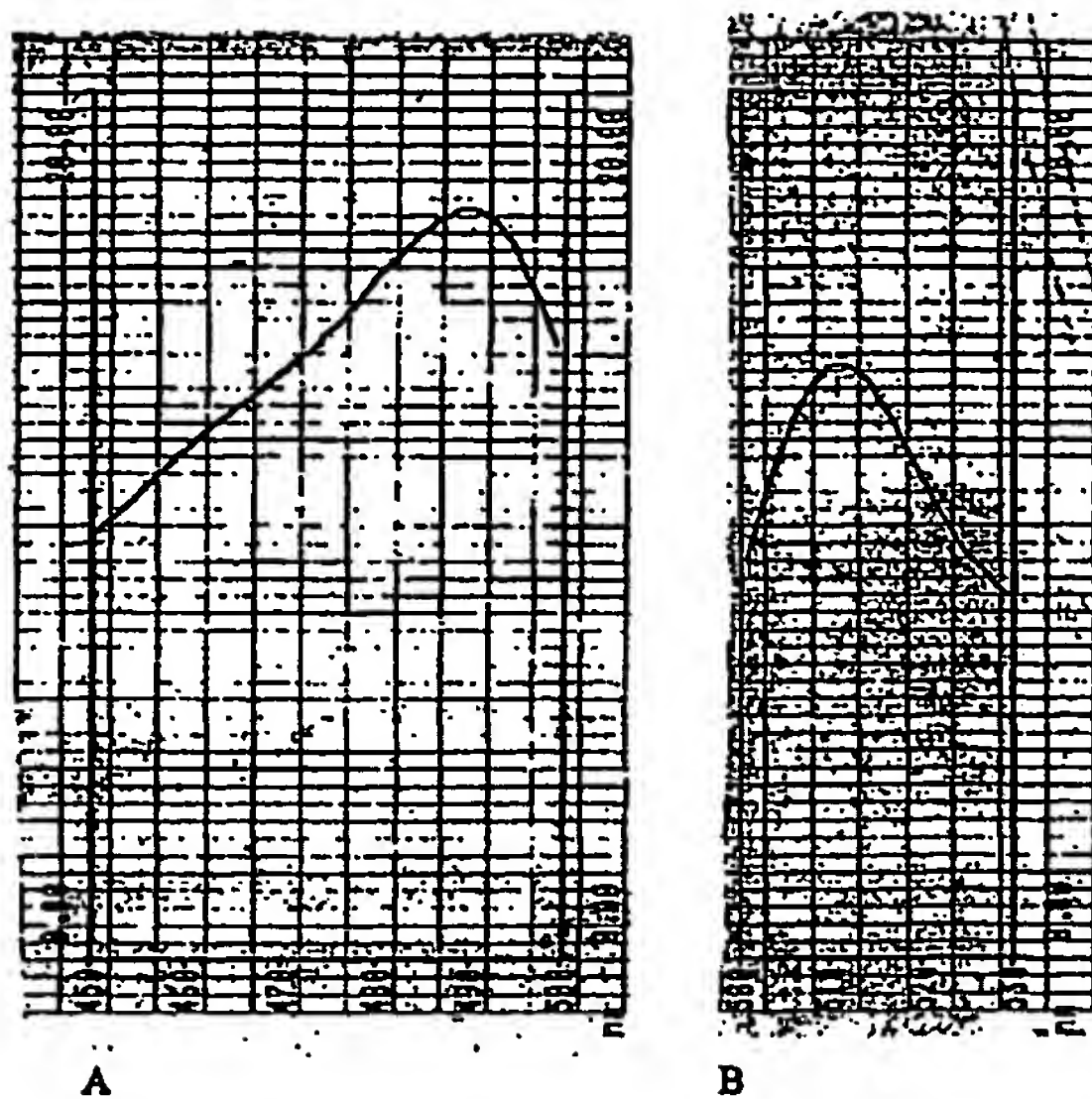


图 8

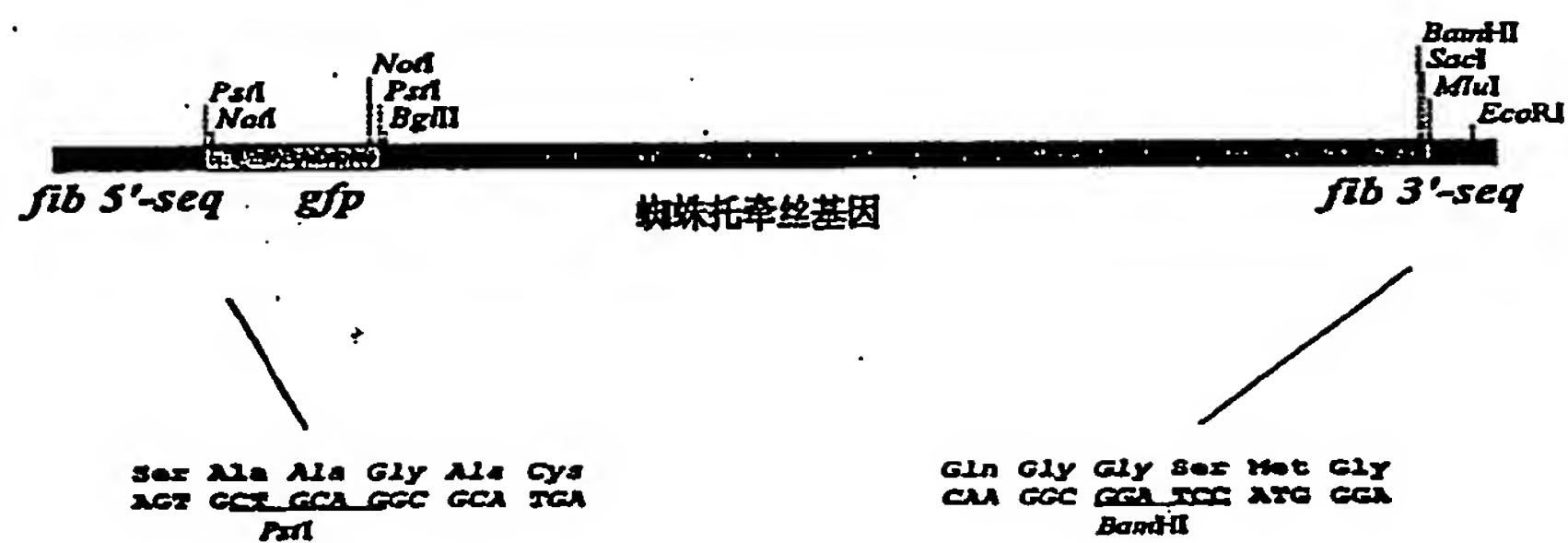


图 9

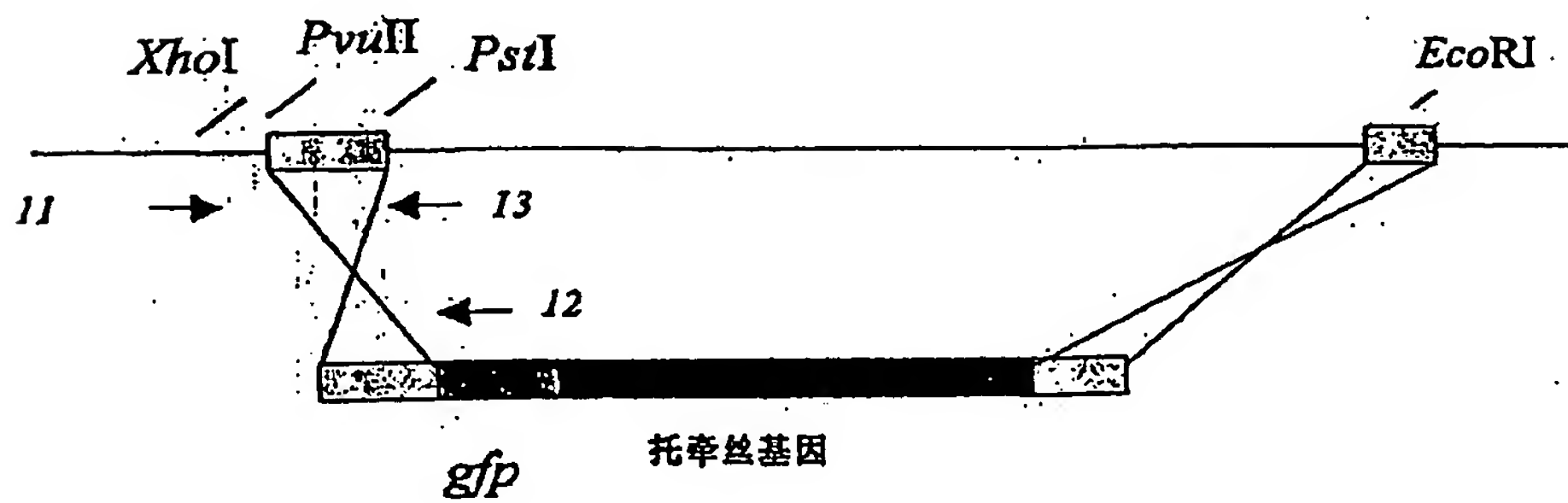


图 11

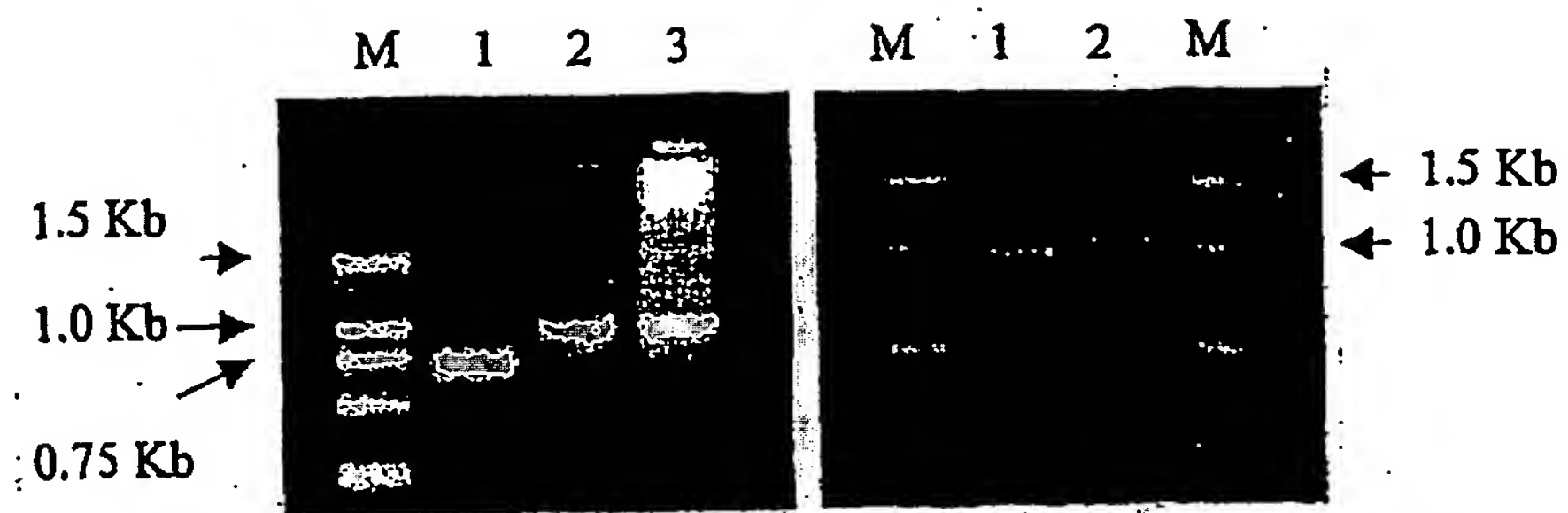


图 12a

图 12b

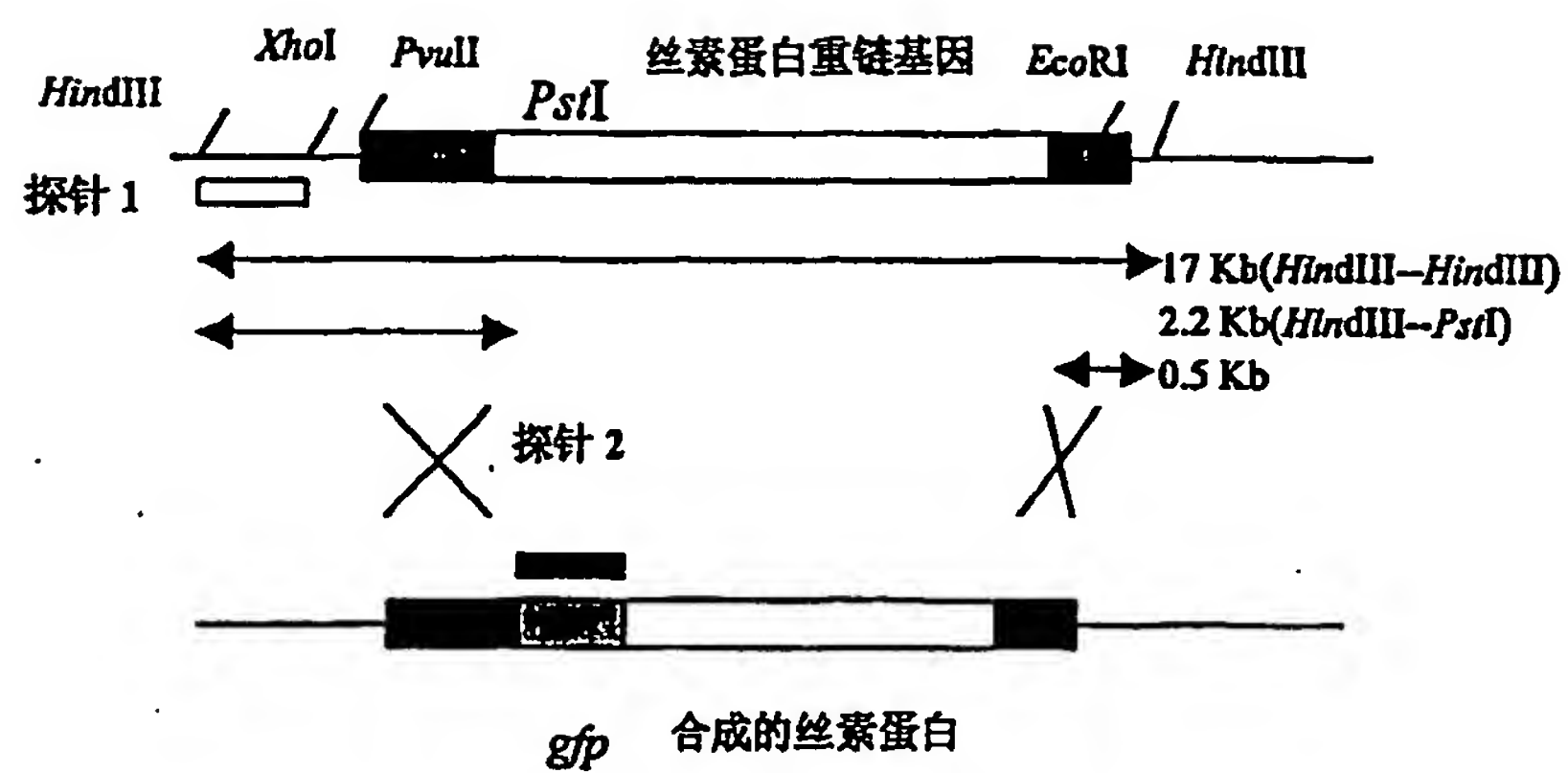


图 13

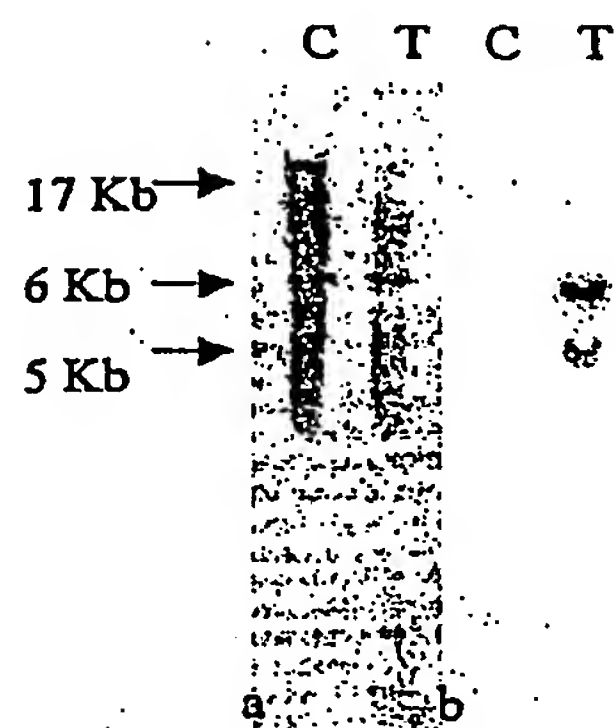


图 14



图 15

5'-ATTAAGCT

TCAAATTCTCTAGAAAAACATT-3' (引物 1)

gtaagtaatttt.....taattattatcaaatctctagaaaaacattgctaagatt.....accgaatttaatttttcgcag

内含子 I

GACGTCATT-5' (引物 2)

3'-GGTAGTTAGTTATGTCACTA

AGCGCCTACGCTGCACCATGGTGACCATCAATCAATACAGTGATAATGAAATTCACGCCACATTGATGATGGAAAAGCTAGG

S A Y A A P S V T I N Q Y S D N E I P R D I D D G K A S

外显子 II

5'-CGTCACAGTGCATCTAGTG-3' (引物 7)

taattaatgtaa.....cgtcacagtgcacatctagtg.....taattattataatttttagt

内含子 II

(引物 3) 5'-TCCTCAACGTTCAAGAGATCTTG-3'

3'-GGGCCACTACGGACACG-5' (引物 8)

TCGGTAATCTCAGTCGATGG.....ATCCTCAACGTTCAAGAGATCTTGAAG.....CCCGGTGATGCCTGTGCAGCCGCTAAC

S V I S R A W I L N V Q E I L K P G D A C A A A N

外显子 III

CTTAAGAT-5' (引物 4)

3'-GGAATTTGTAGGTTATAATGG

gtaagtagatgccgctg.....ccttaaacatccaatattacc.....ttaaaatcttctttgcag

内含子 III

图 16

gtaagtaa.....gttttcgtgaacacggagca.....gaatttaatttttcgcag

(引物 11)→

内含子 I

AGCGCCTACGCTGCACCATCGGTGACCATCAATCAATACAGTGATCTGCAGAGCAAGGGCGAGGAACTGTTCACT
S A Y A A P S V T I N Q Y S D L Q S K G E E L F TGGCGTGGTCCCAATTCTCGTGGAAGTGGATGGCGATGTGAATGGGCACAAATTTTCTGTCAGCGGAGAGGGTGAAG
GTGATGCCACATACGGAAAGCTCACCCCTGAAATTCATCTGCACCACTGGAAAGCTCCCTGTGCCATGGCCAACACT

←(引物 12)

GGTCACTACCTTCACCTATGGCGTGCACTGCTTTTCCAGATACCCAGACCATATGAAGCAGCATGACTTTTTCAAG
AGCGCCATGCCCGAGGGCTATGTGCAGGAGAGAACCATCTTTTCAAAGATGACGGGAACCTACAAGACCCGCGCTG
AAGTCAAGTTCGAAGGTGACACCCTGGTGAATAGAATCGAGCTGAAGGGCATTGACTTTAAGGAGGATGGAACAT
TCTCGGCCACAAGCTCGAATAACAATACTCCACAAATGTGTACATCATGGCCGACAAGCAAAAGATGGCATC
AAGGTCACTTCAAGATCAGACACAACATTGAGGATGGATCCGTGCAGCTGCCCGACCATATCAACAGAACACTC
CAATCGGCGACGGCCCTGTGCTCCTCCAGACAACCATTACCTGTCCACCCAGTCTGCCCTGTCTAAAGATCCCAA
CGAAAGAGAGACCACATGGTCTCTGCTGGAGTTTGTGACCGCTGCTGGGATCACACATGGCATGGACGAGCTGTAC

(引物 9)→

M D E L Y

AAGGAGATCTTGAAGGACATG.....CCCGGTGATGCCTGTGCAGCCGCTAAC

K E I L K D M

P G D A C A A A N

外显子 II—GFP 基因—外显子 III

gtaagtagatgccgctg.....actcgtcatcaaccctg.....ttaaaatcttctttgcag

←(引物 10)

内含子 III

图 17

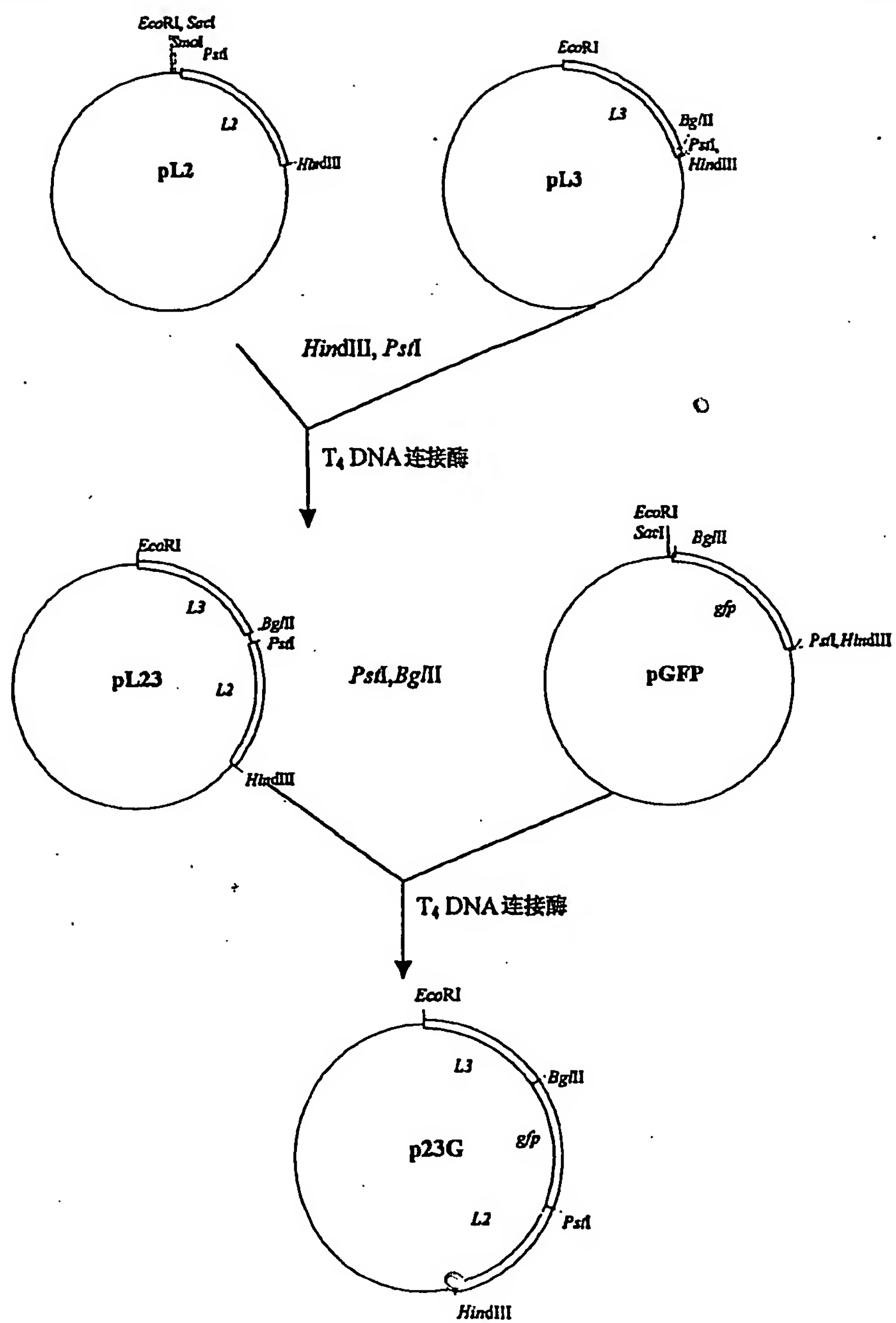


图 18

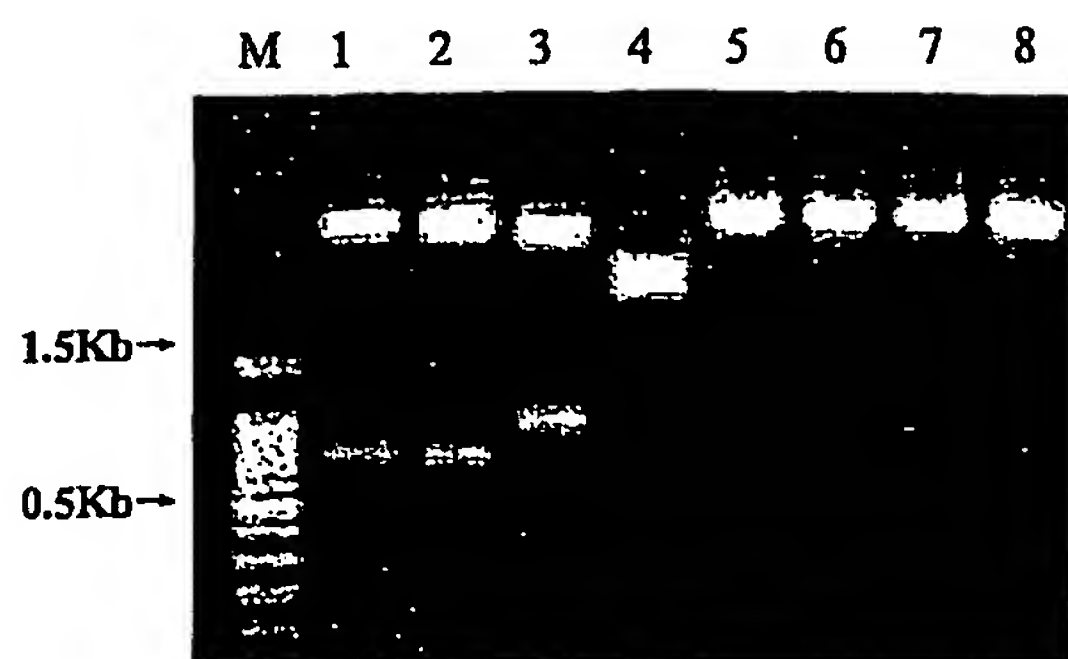


图 19

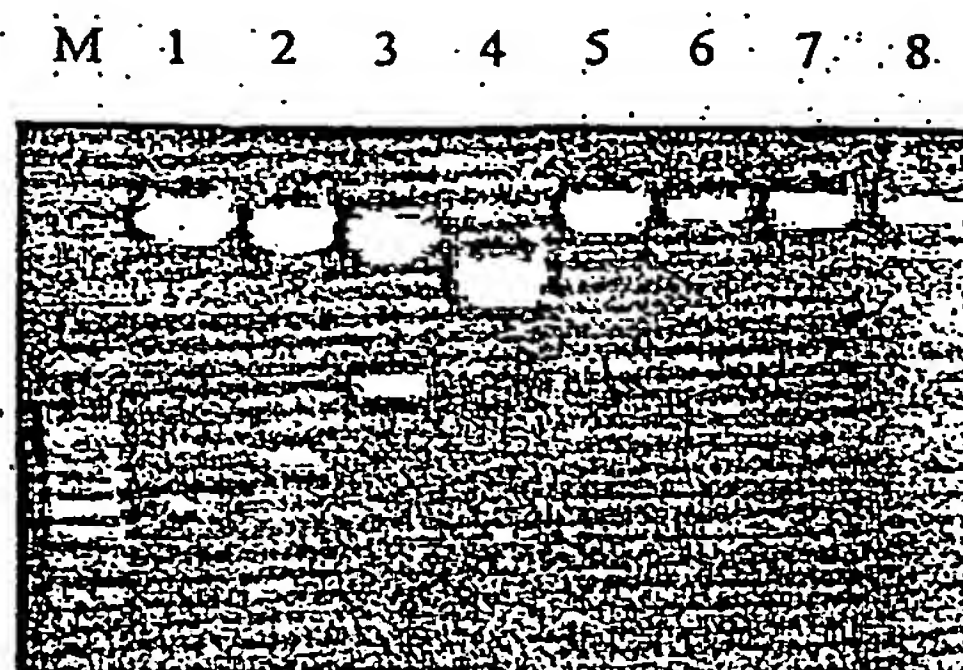


图 25



图 20

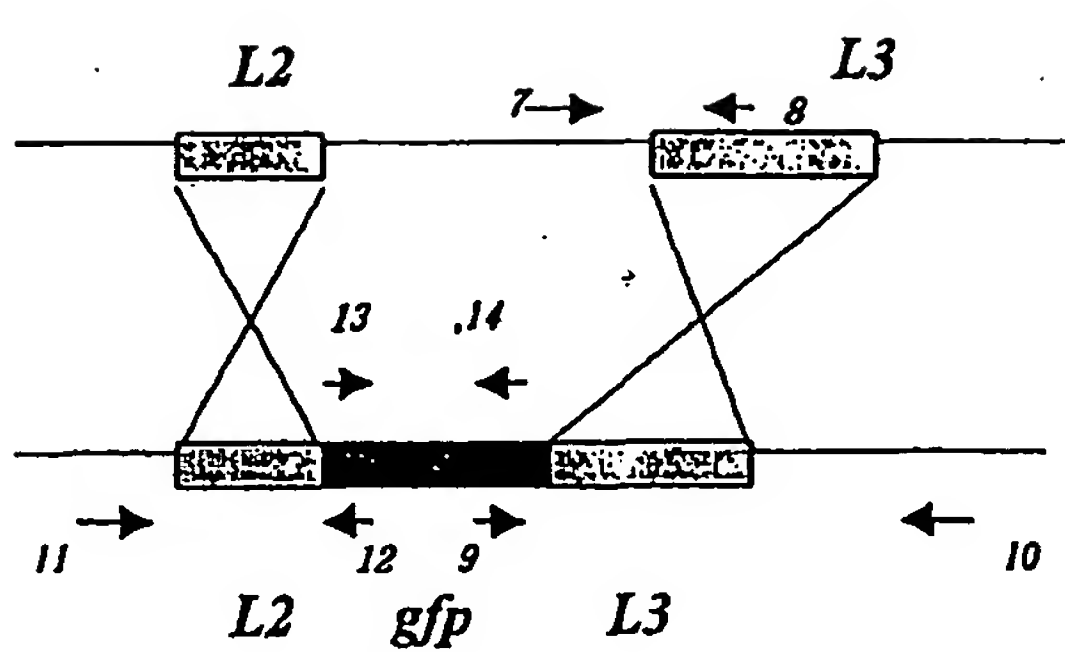


图 21

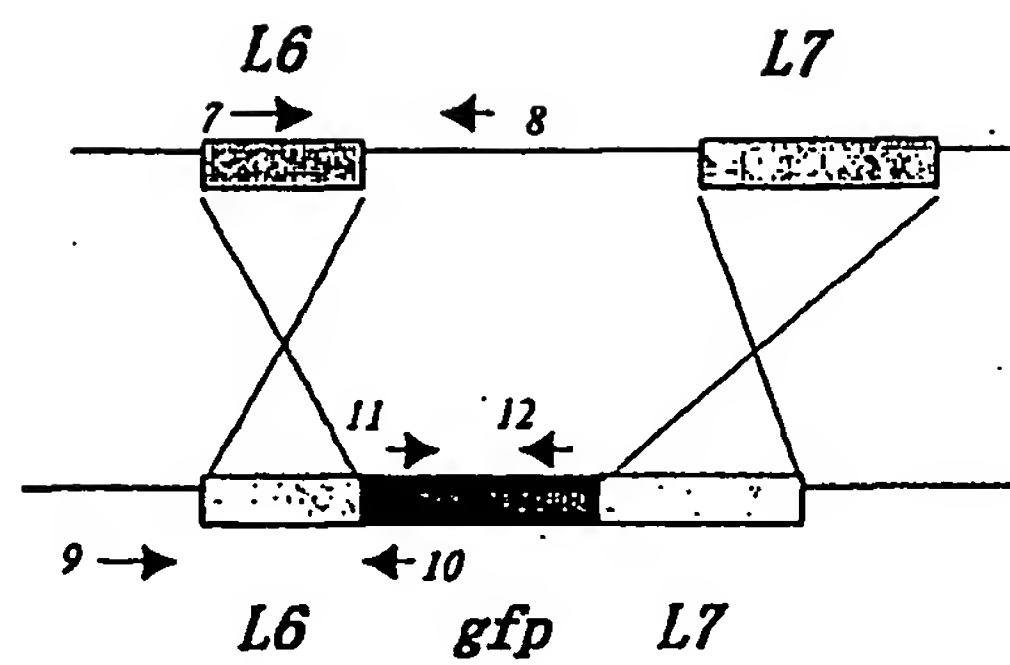


图 26

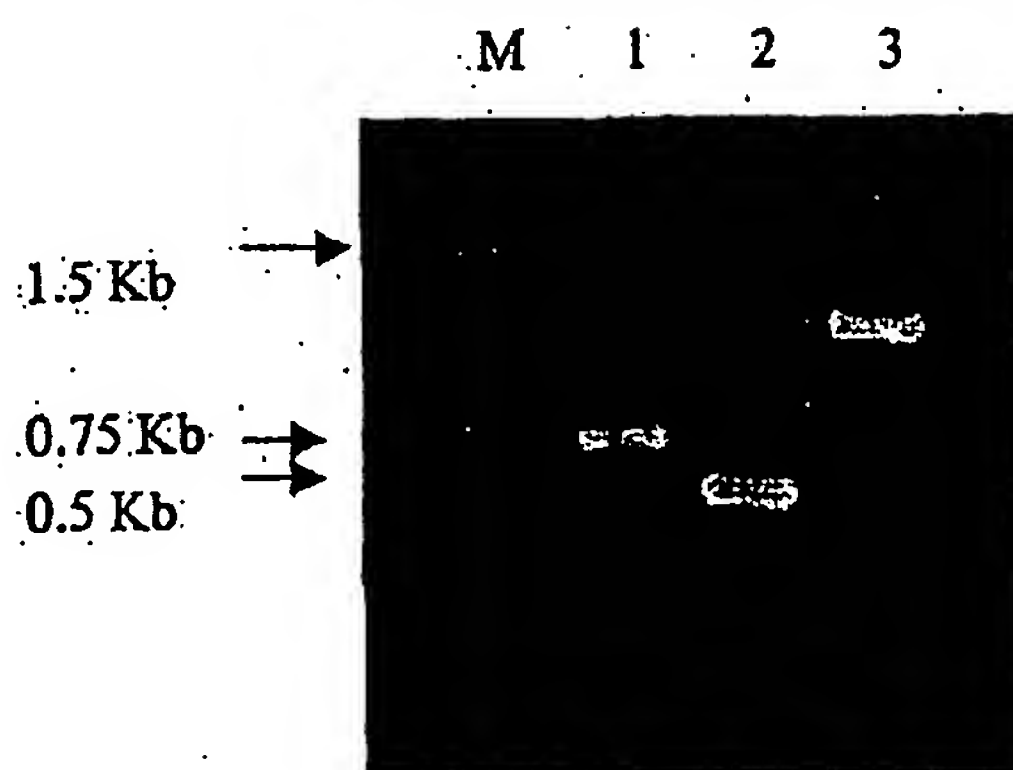


图 27

5'-ATTAA
GCTTTCCAAATAAGAATGTAA-3' (引物 1)
taagcata...gctttccaaataagaatgtaa...cagtcacacagatcggcc...accgaatttaattt
(引物 7) →
内含子 5
GACGTCTAA-5' (引物 2)
3'-GTGGTTAGTCGACCTAGT
TTCTTAAATGAAGAGTAC.....CACCAATCAGCTGGATCAATCACAGAC
F L N E E Y H Q S A G S I T D
外显子 6
gtaagttacg.....ttggacagaaatgagcttacctaattattataatttttagt
← (引物 8)
内含子 6
5'-AGATC
TATATTGCACAAGCAGCCA-3' (引物 3)
CTCCTGAGAGGCGTTGGCAAC.....TATATTGCACAAGCAGCCAGCCAGGTTTCACGTCTAA
L L R G V G N Y I A Q A A S Q V H V ***
外显子 7
CTTAAGTA-5' (引物 4)
3'-TACCTAAAGCCATGACRAC
ataagaactgtaaataaatgtata.....atggatttcggtactgttg.....
3'-非翻译区 (UTR)

图 22

taagcatattgagcttt...gccagcagtgactctagg...attgtattttaccgaatttaatttttcgcag
(引物 9) →
内含子 5
TTCTTAAATGAAGAGTAC.....CACCAATCAGCTGGATCACTGCAGAGCAAGGGCGAGGAACTGTTCACTGGCGTG
F L N E E Y H Q S A G S L Q S K G E E L F T G V
GTCCCAATTCTCGTGGAAGTGGATGGCGATGTGAATGGGCACAAATTTTCTGTCAGCGGAGAGGCTCAAGCTGATG
CCACATACGGAAAGCTCACCCTGAAATTCTCTGCACCACTGGAAAGCTCCCTGTGTCATGGCCAACACTGGTCAC
← (引物 10)
TACCTTCACCTATGCCGTGCAGTGCCTTTTCCAGATACCCAGACCATATGAAGCAGCATGACTTTTTCAAGAGCGCC
ATGCCCGAGGGCTATGTGCAGGAGAGAACCATCTTTTCAAGATGACCGGAACTACAAGACCCGCGCTGAAGTCA
AGTTTGAAGGTGACACCCCTGGTGAATAGAATCGAGCTGAAGGGCATTCACTTTAAGGAGGATGGAACATTCTCGG
CCACAAGCTGGAATACAACCTATACTCCCAATGTGTACATCATGGCCGACRAGCAAAAGAATGGCATCAAGGTC
AACTTCAAGATCAGACACAACATTGAGGATGGATCCGTGCAGCTGGCCGACCATTATCAACAGAACACTCCAATCG
CGGACGGCCCTGTGCTCCTCCCAGACAACCATTACCTGTCCACCCAGTCTGCCCTGTCTAAAGATCCCAACGAAA
GAGAGACCACATGCTCCTGCTGGAGTTTGTGACCGCTGCTGGGATCACACATGGCATGGACGAGCTGTACAAG
M D E L Y K
GAGATCTTATATTGCACAAGCAGCCAGCCAGGTTTCACGTCTAA
E I Y I A Q A A S Q V H V ***

外显子 6-GFG 基因-外显子 7

图 23

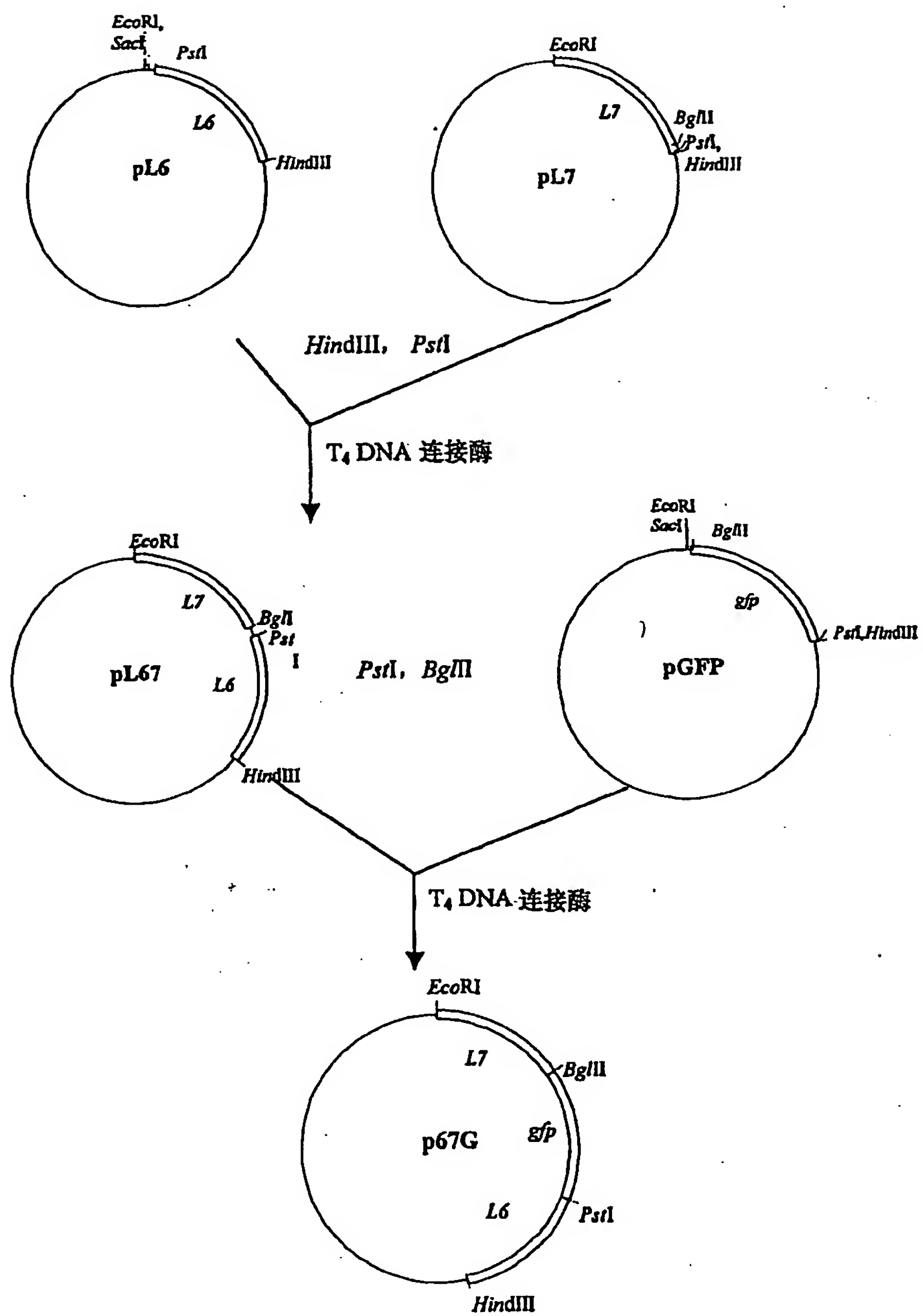


图 24

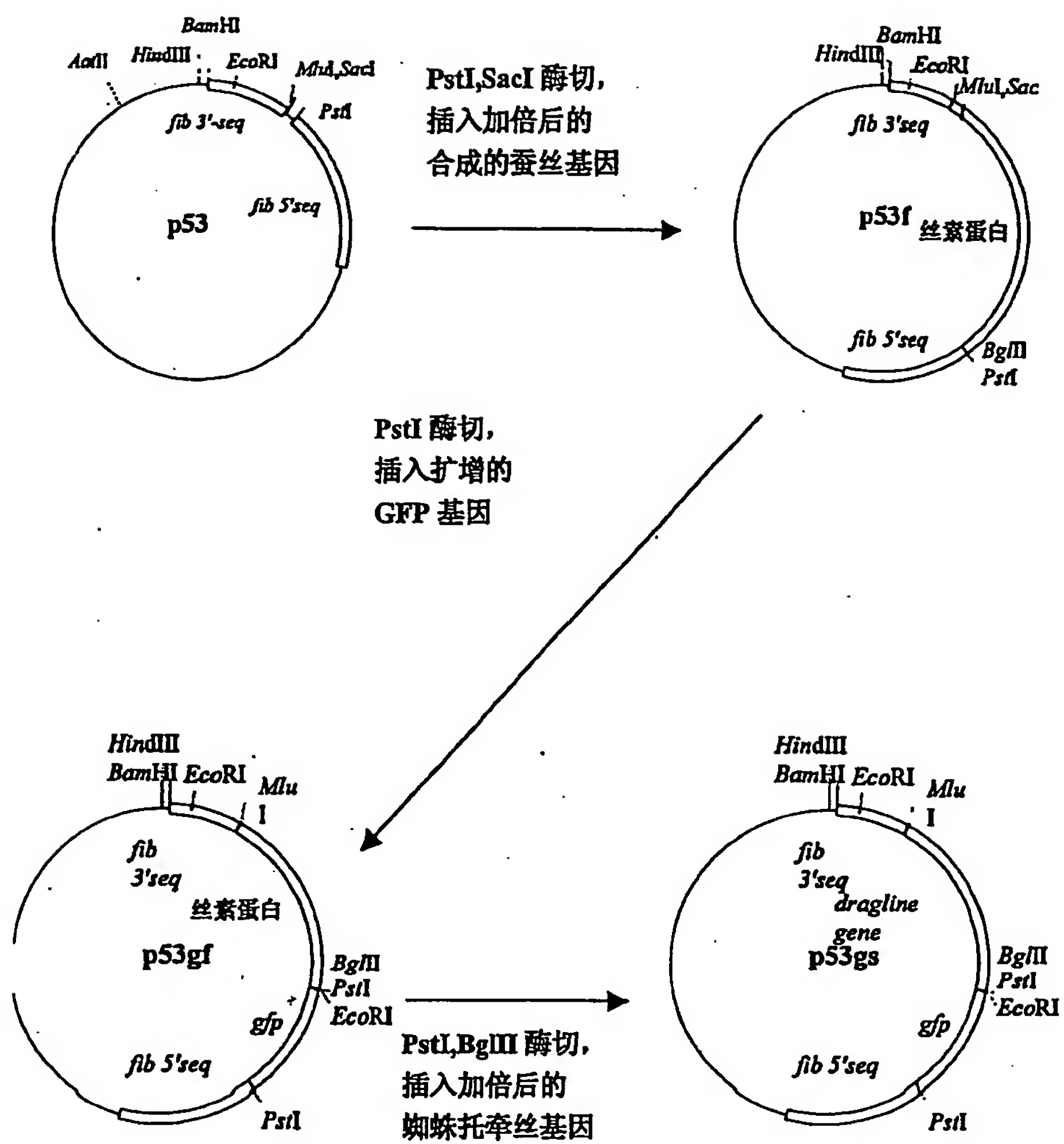


图 28

序列表

<110> 中国科学院上海生物化学研究所

陆, 长德
黄, 君霆
赵, 昀
张, 峰
陈, 秀

<120> 利用家蚕生产非天然蚕丝的方法

<130> 003834 PCWO

<150> CN 00125859.1

<151> 2000-10-26

<160> 49

<170> PatentIn version 3.0

<210> 1

<211> 655

<212> PRT

<213> 人工序列

<220>

<221> misc_feature

<223> 蜘蛛拖牵丝蛋白-绿色荧光蛋白融合蛋白

<400> 1

Met Ala Ala Gly Gly Met Ser Lys Gly Glu Glu Leu Phe Thr Gly Val
1 5 10 15

Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe
20 25 30

Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr
35 40 45

Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr
50 55 60

Leu Val Thr Thr Phe Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro
65 70 75 80

Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly
85 90 95

Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys
100 105 110

Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile
115 120 125

Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His
130 135 140

Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp

145	150	155	160
Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile			
165	170	175	
Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro			
180	185	190	
Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr			
195	200	205	
Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val			
210	215	220	
Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu			
225	230	235	240
Leu Tyr Lys Gly Ile Pro Ala Asp Gly Arg Ser Gly Gly Leu Gly Gly			
245	250	255	
Gln Gly Pro Gly Gln Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala			
260	265	270	
Gln Gly Gly Thr Gly Gly Leu Gly Ser Gln Gly Ala Gly Arg Gly Gly			
275	280	285	
Gln Gly Ala Ala Ala Ala Ala Ala Val Glu Gly Gly Ala Gly Gln Gly			
290	295	300	
Gly Thr Gly Gly Leu Gly Ser Gln Gly Ala Gly Arg Gly Gly Leu Gly			
305	310	315	320
Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala Gly			
325	330	335	
Gln Gly Gly Thr Gly Gly Pro Gly Asn Gln Gly Gly Ser Met Gly Ser			
340	345	350	
Gly Gly Leu Gly Gly Gln Gly Pro Gly Gln Gly Ala Ala Ala Ala Ala			
355	360	365	
Ala Ala Gly Gly Ala Gln Gly Gly Thr Gly Gly Leu Gly Ser Gln Gly			
370	375	380	
Ala Gly Arg Gly Gly Gln Gly Ala Ala Ala Ala Ala Val Glu Gly			
385	390	395	400
Gly Ala Gly Gln Gly Gly Thr Gly Gly Leu Gly Ser Gln Gly Ala Gly			
405	410	415	
Arg Gly Gly Leu Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala			
420	425	430	
Ala Gly Gly Ala Gly Gln Gly Gly Thr Gly Gly Pro Gly Asn Gln Gly			
435	440	445	
Gly Ser Met Gly Ser Gly Gly Leu Gly Gly Gln Gly Pro Gly Gln Gly			
450	455	460	
Ala Ala Ala Ala Ala Ala Ala Gly Gly Ala Gln Gly Gly Thr Gly Gly			

465 470 475 480
 Leu Gly Ser Gln Gly Ala Gly Arg Gly Gly Gln Gly Ala Ala Ala Ala
 485 490 495
 Ala Ala Val Glu Gly Gly Ala Gly Gln Gly Gly Thr Gly Gly Leu Gly
 500 505 510
 Ser Gln Gly Ala Gly Arg Gly Gly Leu Gly Gly Gln Gly Ala Gly Ala
 515 520 525
 Ala Ala Ala Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly Thr Gly Gly
 530 535 540
 Pro Gly Asn Gln Gly Gly Ser Met Gly Ser Gly Gly Leu Gly Gly Gln
 545 550 555 560
 Gly Pro Gly Gln Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala Gln
 565 570 575
 Gly Gly Thr Gly Gly Leu Gly Ser Gln Gly Ala Gly Arg Gly Gly Gln
 580 585 590
 Gly Ala Ala Ala Ala Ala Ala Val Glu Gly Gly Ala Gly Gln Gly Gly
 595 600 605
 Thr Gly Gly Leu Gly Ser Gln Gly Ala Gly Arg Gly Gly Leu Gly Gly
 610 615 620
 Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala Gly Gln
 625 630 635 640
 Gly Gly Thr Gly Gly Pro Gly Asn Gln Gly Gly Ser Met Gly Ala
 645 650 655

<210> 2
 <211> 38
 <212> DNA
 <213> 引物

<400> 2
 atagaattcc tgcaggcggc atgagcaagg gcgaggaa

38

<210> 3
 <211> 37
 <212> DNA
 <213> 引物

<400> 3
 aaactgcagg aattcccttg tacagctcgt ccatgcc

37

<210> 4
 <211> 39
 <212> DNA
 <213> 接头

<400> 4
 taatagact cactataggg aagtctcgag acttccta

39

<210> 5

<211> 47
<212> DNA
<213> 接头

<400> 5
agcttaggga agtctcgaga cttccctata gtgagtcgta ttatgca

47

<210> 6
<211> 33
<212> DNA
<213> 接头

<400> 6
agctttatat ttggaggtaa tcatatggct gca

33

<210> 7
<211> 25
<212> DNA
<213> 接头

<400> 7
gccatgatgat tacctccaaa tataa

25

<210> 8
<211> 29
<212> DNA
<213> 接头

<400> 8
catcaccatc accatcacta atgctagcg

29

<210> 9
<211> 37
<212> DNA
<213> 接头

<400> 9
aattcgctag cattagtgat ggtgatggtg atgagct

37

<210> 10
<211> 23
<212> DNA
<213> 引物

<400> 10
tagaattcag aagtgaaacg ccg

23

<210> 11
<211> 28
<212> DNA
<213> 引物

<400> 11
taagacgtcg atggcttgta gatatgac

28

<210> 12
<211> 16
<212> DNA
<213> 接头

<400> 12
agcttctcga gctgca

16

<210> 13
<211> 16
<212> DNA
<213> 接头

<400> 13
agagctcggc tcgaga

16

<210> 14
<211> 8
<212> DNA
<213> 接头

<400> 14
cacgcgtg

8

<210> 15
<211> 16
<212> DNA
<213> 接头

<400> 15
aattcacgcg tgagct

16

<210> 16
<211> 18
<212> DNA
<213> 引物

<400> 16
attttcataa ccctcgag

18

<210> 17
<211> 17
<212> DNA
<213> 引物

<400> 17
accgactgca gcactag

17

<210> 18
<211> 26
<212> DNA
<213> 引物

<400> 18
taaacgcgtg actttgaaac tggaag

26

<210> 19
<211> 17
<212> DNA
<213> 引物

<400> 19
ttgtcttcta ggaattc

17

<210> 20

<211> 21
<212> DNA
<213> 引物

<400> 20
ttttcataac ctcgaggtag c

21

<210> 21
<211> 19
<212> DNA
<213> 引物

<400> 21
catggcacag ggagctttc

19

<210> 22
<211> 21
<212> DNA
<213> 引物

<400> 22
ttttcataac ctcgaggtag c

21

<210> 23
<211> 18
<212> DNA
<213> 引物

<400> 23
ctcctgcacc gactgcag

18

<210> 24
<211> 17
<212> DNA
<213> 引物

<400> 24
atgagcaagg gcgagga

17

<210> 25
<211> 18
<212> DNA
<213> 引物

<400> 25
ttgtacagct cgtccatg

18

<210> 26
<211> 23
<212> DNA
<213> 引物

<400> 26
tataagcttg ttgtacaaaa ctg

23

<210> 27
<211> 18
<212> DNA
<213> 引物

<400> 27
tgctacctcg aggttatg

18

<210> 28
<211> 20
<212> DNA
<213> 引物

<400> 28
gccagtgaat tcctagaaga

20

<210> 29
<211> 19
<212> DNA
<213> 引物

<400> 29
ccaggacgaa gtaagaaac

19

<210> 30
<211> 30
<212> DNA
<213> 引物

<400> 30
attaagcttc aaattctcta gaaaaacatt

30

<210> 31
<211> 29
<212> DNA
<213> 引物

<400> 31
ttactgcaga tcactgtatt gattgatgg

29

<210> 32
<211> 20
<212> DNA
<213> 引物

<400> 32
tcaacgttca agagatcttg

20

<210> 33
<211> 29
<212> DNA
<213> 引物

<400> 33
tagaattcgg taatattgga tgtttaagg

29

<210> 34
<211> 26
<212> DNA
<213> 引物

<400> 34
atactgcaga gcaagggcga ggaact

26

<210> 35

<211> 28
<212> DNA
<213> 引物

<400> 35
attagatctc cttgtacagc tcgtccat

28

<210> 36
<211> 19
<212> DNA
<213> 引物

<400> 36
cgtcacagtg catctagt

19

<210> 37
<211> 17
<212> DNA
<213> 引物

<400> 37
gcacaggcat caccggg

17

<210> 38
<211> 19
<212> DNA
<213> 引物

<400> 38
cctgctggag tttgtgacc

19

<210> 39
<211> 17
<212> DNA
<213> 引物

<400> 39
cagggttgat gacgagt

17

<210> 40
<211> 20
<212> DNA
<213> 引物

<400> 40
gttttcgtga acacggacga

20

<210> 41
<211> 19
<212> DNA
<213> 引物

<400> 41
catggcacag ggagctttc

19

<210> 42
<211> 27
<212> DNA
<213> 引物

<400> 42
aatctgcagt gatccagctg attggtg

27

<210> 43
<211> 26
<212> DNA
<213> 引物

<400> 43
attaagcttt ccaaataaga atgtaa

26

<210> 44
<211> 24
<212> DNA
<213> 引物

<400> 44
agatctatat tgcacaagca gccca

24

<210> 45
<211> 27
<212> DNA
<213> 引物

<400> 45
atgaattcca acagtaccga aatccat

27

<210> 46
<211> 18
<212> DNA
<213> 引物

<400> 46
cagtcaacag atatcgcc

18

<210> 47
<211> 19
<212> DNA
<213> 引物

<400> 47
gataagctca tctgtccaa

19

<210> 48
<211> 18
<212> DNA
<213> 引物

<400> 48
gccagcagtg actctagg

18

<210> 49
<211> 19
<212> DNA
<213> 引物

<400> 49
catggcacag ggagctttc

19

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN01/01506

A. CLASSIFICATION OF SUBJECT MATTER

IPC7 C07K14/435, C12N15/12, C12N15/85, C12N5/16

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

GenBank WPI CNPAT

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Proc. Natl. Acad. Sci. U.S.A. 93 (10), 4845-4850 (1996) Htun, H., et al., Visualization of glucocorticoid receptor translocation and intranuclear organization in living cells with a green fluorescent protein chimera	1-13
A	Mol. Biol. Cell 10 (2), 471-486 (1999) Htun, H., et al., Direct visualization of the human estrogen receptor alpha reveals a role for ligand in the nuclear distribution of the receptor	1-13

Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
1 Mar. 2002 (01.03.02)

Date of mailing of the international search report
04. April 2002 104.04.02

Name and mailing address of the ISA/CN
6 Xitucheng Rd., Jimen Bridge, Haidian District,
100088 Beijing, China
Facsimile No. 86-10-62019451

Authorized officer

JIA

Telephone No. 86-10-62093933



国际检索报告

国际申请号

PCT/CN01/01506

A. 主题的分类

IPC7

C07K14/435,

C12N15/12,

C12N15/85, C12N5/16

按照国际专利分类表(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类体系和分类号)

IPC7

C07K

C12N

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称和, 如果实际可行的, 使用的检索词)

GenBank WPI CNPAT

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求编号
A	Proc. Natl. Acad. Sci. U.S.A. 93 (10), 4845-4850 (1996), Htun,H.等., Visualization of glucocorticoid receptor translocation and intranuclear organization in living cells with a green fluorescent protein chimera	1-13
A	Mol. Biol. Cell 10 (2), 471-486 (1999) Htun,H.,等, Direct visualization of the human estrogen receptor alpha reveals a role for ligand in the nuclear distribution of the receptor	1-13

其余文件在C栏的续页中列出。

☒ 见同族专利附件。

* 引用文件的专用类型:

"A" 明确叙述了被认为不是特别相关的一般现有技术的文件

"B" 在国际申请日的当天或之后公布的在先的专利或专利申请

"L" 可能引起对优先权要求的怀疑的文件, 为确定另一篇

引用文件的公布日而引用的或者因其他特殊理由而引用的文件

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布的在后文件, 它与申请不相

抵触, 但是引用它是为了理解构成发明基础的理论或原理

"X" 特别相关的文件, 仅仅考虑该文件, 权利要求所记载的

发明就不能认为是新颖的或不能认为是有创造性

"Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件

结合并且这种结合对于本领域技术人员为显而易见时, 权利要求记载的发明不具有创造性

"&" 同族专利成员的文件

国际检索实际完成的日期

1.3 月 2002(01.03.02)

国际检索报告邮寄日期

04. 4月 2002 (04. 04. 02)

国际检索单位名称和邮寄地址

ISA/CN

中国北京市海淀区西土城路6号(100088)

传真号: 86-10-62019451

受权官员



电话号码: 86-10-62093933